THE MAJESTY OF REASON

A Short Guide to Critical Thinking in Philosophy

JOSEPH C. SCHMID

The Majesty of Reason

A Short Guide to Critical Thinking in Philosophy

Joseph C. Schmid

"What altar of refuge can a man find for himself when he commits treason against the majesty of reason?"

– Baruch Spinoza

"Reason is immortal; all else mortal."

– Pythagoras

Chains

That which surrounds is chained – both mind and flesh.
Be self-imposed, or maybe not? We bow
To glory, pride, repose without refresh;
Our chains we slaves obsequiously avow.
We wallow in a sea of filth, not truth;
Proclaim adherence to repeated breath.
Our only hope? Corruption of the youth,
The persecuted spirit – put to death.
Corruption is enlightening! The doubt
That soaks accustomed minds drips freedom, awe!
Relinquish bondage, leave the cave, throw out
The rotten, adore Her beauty, ditch straw
for steel! The thought, the thought, the reasoned thought.
The end result? Our chains reduced to naught.

Contents

Chapter 1: Setting the Stage

1 Introduction

Polarization ravages our current ideological climate. The battle of ideas rages on among disagreeing factions of society. Warring parties, proudly waving their flags of allegiance, gear up for ideological battle. Such tribalism festers in countless domains of inquiry. But what is tribalism? And what is the antidote to this division?

Tribalism is the dogmatic and weaponized adherence to the principles and beliefs of a social, political, or ideological group. In the battle of ideas, tribalism fosters us-versus-them mentality, breeds confirmation bias, cultivates prideful ignorance, and promotes the devaluation of people outside the tribe. In tribalistic thinking, arguments are used as weapons meant to attack rather than tools designed to serve; interlocutors are stripped of their value, antagonized, and presupposed irrational.

Tribalism underpins so many of the problems plaguing humanity. The devaluation and systematic murder of religious, social, and racial groups in parts of the world is fueled by the dehumanization of those outside one's ideological tribe. Tribalism is no trivial matter; its pernicious effects are witnessed on scales ranging from interpersonal communication to systematic genocide.

There is no doubt that the ingrained shortcomings of human nature are partially responsible for this tribalism. We naturally tend to seek out information that supports our pre-existing beliefs and revel in the thought of crushing an opponent in a heated debate. We have lost sight of the very purpose for which dialogue and discussion exist in the first place, the very treasure upon which we base our intellectual pursuits: truth. We have forgone our duty to make truth our ultimate aim and, in doing so, have allowed tribalism to germinate, sewing its deleterious effects into the very fabric of our thoughts, worldviews, and discussions.

Without a recognition that truth is our aim, we see interlocutors not as fellow explorers on a journey towards a common treasure, but as villains with whom we must compete to gain status, pride, and ego-gratification. We employ polemics and debate tactics to conquer our opponents rather than instilling a dialectical foundation of respect, mutual dialogue, and genuine curiosity. We dehumanize our opponents and assume their close-mindedness, irrationality, and ignorance.

Rather than fundamentally alter human nature, then, as an antidote to this dire ideological situation, the solution is a mutual recognition that we are fellow explorers who can *learn* from one another and seek the beautiful treasure of truth *together* as an enterprise of curiosity, respect, and love. It is only through valuing and loving others, encouraging critical and reflective thought, and aiming ourselves toward truth that we can solve the polarizing problems afflicting humanity.

But merely to identify the antidote to tribalism doesn't take us very far. Without developing the skills and tools of critical and reflective thought, without actually orienting ourselves toward truth, without seeing each other as fellow explorers, we are left chained to the very thing for which we sought a cure. Identifying the antidote without tangible action towards its realization makes us cave dwellers who, after discovering the path to the illumined outside world, embrace the cold and darkened depths in which we have toiled. Instead of embracing the shackles of ignorance and the bondage of tribalism, then, we must seek liberation. We must, to use Thomas Jefferson's famous injunction, "preach a crusade against ignorance."

How do we, as explorers, as seekers, proceed on this journey? How do we follow the path towards illumination rather than embrace the darkness of tribalism? How do we discover the treasure of truth?

It is precisely these questions that this book aims to address. The first step on our journey lays the foundation for the ensuing steps. It equips you with the intellectual virtues and conversational tactics necessary for critical thinking. Attitudes and dispositions are just as important to critical thinking

as being well-versed in fallacies, argumentation, and so on. Without cultivating the dispositions of curiosity, humility, open-mindedness, a love for truth, and more, a proficiency in argumentation is groundless and without direction. This foundation will occupy the remainder of the present chapter (save the next few paragraphs, of course).

The second step on our journey equips you with a variety of methods and tools in critical thinking and philosophical reasoning. Through tangible suggestions, lively and engaging examples, and a bit of technical jargon, you'll come away a better thinker and – ideally – a better human being.

Critical thinking, like mastering a musical instrument, requires practice. That's why the next two steps on our journey will apply a number of the methods and tools previously explored to central issues in philosophy of science and philosophy of mind: scientism, laws of nature, and the nature of the mind. Don't worry if you're unfamiliar with these issues – they will be explained in due time.

Why these issues, though? Well, everyone loves science to *some* degree. Whether it's volcanoes, engines, pterodactyls, wave functions, the origin of the universe, the psychology of social media, or black holes, humans are *empirical* creatures. Given this universal interest, what better place exists to embark on a journey in the arts of critical thinking and philosophical reasoning? What's more, consciousness is one of the most intimate yet mysterious phenomena in the world. The puzzles it raises are fascinating and, most importantly, bear on the fundamental truths of human nature.

This book is for, well, *everyone*. If you are unaccustomed to philosophy, it might stretch you at times. But don't fret – that's a *good* thing. If you possess substantial, expert-level knowledge in philosophy, there is plenty in here that will be of benefit to you. Novel arguments pertaining to philosophy of science, metaphysics, and philosophy of mind are developed and evaluated within later chapters, and the methods and tools of critical thinking will be helpful to philosophers no less than philosophy students.

So, we've covered (i) the importance of this book, (ii) the general structure of our journey, and (iii) for whom this book is intended.[1] One last thing before we delve into the juicy stuff. The ultimate aim of this book is to *inspire*. It is to inspire your curiosity, your wonder, your hunger for knowledge. It is to inspire an impassioned and flaming love for truth, an appetite for learning, and an appreciation for the majesty of reason.

We are explorers, seekers. With the tools of rational and critical thinking in hand, we embark on our journey towards truth, eschewing tribalism and us-versus-them thinking.

In my journey, I am sometimes engulfed by the beauty and majesty of reason, its immortality, its power to peer into the depths of reality, its pure and unbridled and unforgivingly sharp logicality, its ability to give unadulterated deductive certitude. I likewise hope to instill in *you* this awe for the majesty of reason. The journey is yours. Never stop wondering. Never let your curiosity be extinguished. The next chapters are just the beginning of your journey of discovery. The treasure of truth awaits.

2 Getting virtuous

The foundation of critical thinking is virtue. Virtues are (i) stable traits or characteristics that (ii) one develops through acts, habits, and character formation with (iii) some intrinsically valuable *telos* (end, purpose, or goal). A morally virtuous person consistently strives for the *telos* of moral perfection through continued action and habituation towards doing good and avoiding evil. An intellectually virtuous person consistently strives for truth and greater understanding through various habits of the mind – intellectual humility, intellectual perseverance, and so on.

Here is a brief survey of some of the most important intellectual virtues that will serve as the foundation for the rest of the book (and, hopefully, for the rest of your life).

Intellectual humility

Being intellectually humble involves recognizing the limitations of your knowledge and abilities. It also involves orienting one's focus on *truth* rather than one's *status* in relation to others. Intellectually humble individuals properly attend to their own weaknesses, but they also don't ignore their strengths. Fundamentally, intellectual humility is the willingness to say 'I don't know,' 'I haven't researched this in sufficient depth,' and 'I might be wrong' in the appropriate circumstances.

Intellectual curiosity

Intellectual curiosity refers to the flaming passion and interest one has for discovering truths in greater breadth and depth. Intellectually curious individuals despise intellectual laziness – the preference for believing what one has always believed because it's comfortable, reassuring, or what have you. Curious folks dare to pursue issues further and deeper, giving such individuals a greater understanding of the complexities and foundations of the topic at hand. Ask questions. Dare to discover more.

Intellectual perseverance

Like curiosity, intellectual perseverance is a committed rejection of intellectual laziness. Those with intellectual perseverance see barriers, hardships, and obstacles in their intellectual life as *challenges to be overcome* and *opportunities for growth*. Instead of giving up their pursuit of truth when it gets tough, they dig even deeper into the issues, read and investigate further, and grow in knowledge and understanding. Problems and puzzles are opportunities for discovery, not insurmountable barriers. Persist in the path of discovery, because treasures await.

Intellectual responsibility

Developing intellectual responsibility is a matter of taking charge of your pursuit of truth. Intellectually responsible individuals try to minimize deferring their beliefs to the hands of authority figures; instead, they seek to equip *themselves* with justification, arguments, evidence, and reasons.

Open-mindedness

Being open-minded is not a matter of believing that every single alternative position is equally probable. On the contrary, it is the willingness to *examine* such alternative positions in order to critically evaluate their rational and evidential merit. Being open-minded is all about a willingness to *consider* alternative viewpoints and to recognize, where appropriate, that one's own position might be mistaken.

Love of truth

We are all susceptible to confirmation bias – the tendency to selectively attend to sources, ideas, and evidence that *support* or *confirm* our pre-existing beliefs and ignore or downplay sources, ideas, and evidence that contravene them. But a genuine love – dare I say *adoration* – for truth can help combat confirmation bias. Don't make *defending a position* your primary aim in intellectual pursuits; instead, make *discovering truth* your foremost goal.

3 Tips for productive conversations

Intellectual virtues are, once again, the foundation of critical thinking. However, they provide little guidance on how to have productive, fruitful, truth-oriented conversations with your dialectical partners (i.e. those with whom you converse). Intellectual virtues are all about directing oneself – *inwardly* – to the *telos* of truth. But conversations are a step beyond that, since they implicate *others* in the pursuit of truth. Intellectual virtues are about action; conversations are about *interaction*.

It stands to reason, then, that the intellectual virtues will be limited in their utility when it comes to conversations. For this reason, I've compiled a number of tips I (along with other philosophers, social psychologists, etc.) have found particularly useful in facilitating the productive exchange of ideas.[2]

Tip #1

Recognize that we are all biased in favor of our current views. It's time to stop thinking of that as a personality flaw; it's a flaw of *human nature*.

Being open minded is *really hard*, and we should appreciate that. Work to mitigate such bias, yes; but don't degrade or look down on others who suffer from it.

Tip #2

Emphasize *agreement*. Too often we get bogged down in our disagreements, failing to see miles upon miles of common ground.

Tip #3

Explicitly recognize that you and your dialectical partners are *on the same team*, with the same goals: gaining a greater understanding of an issue and discovering treasures of truth. It's a mutual effort. We are on the same path towards a common destination. *Explorers on a journey*.

Tip #4

Be willing to *learn* from your dialectical partner. Everyone has unique experiences, perspectives, and worldviews from which we can gain value and insights. Don't go into the conversation with the sole aim of trying to convert or convince them.

Tip #5

Say 'I don't know' when you genuinely don't know. There is beauty in admitting that further paths of discovery and treasure await you.

Tip #6

Curb certainty. It's okay (indeed, good!) to be certain on some positions, but it should be somewhat rare. Certainty can bleed into dogmatism, and that is to be avoided. Recognize that you might be wrong.

Tip #7

We can't rid our biases, but we *can* modify dialectical contexts to mitigate them. Create a welcoming, respectful environment where truth and love are center-stage.[3]

Tip #8

Perspective-take. Think about the life experiences of your dialectical partner, and place yourself in his or her shoes. This helps to *humanize* them rather than dehumanize them.

Tip #9

Ditch the caricatures. We are *individuals* with loves, passions, desires, and value. No one fits a pre-defined caricature or stereotype. We have our unique quirks, flaws, and skills. Consciously try to break pre-existent caricatures of the 'type' of person with whom you're speaking.

Tip #10

Immersion. Immerse yourself on a daily basis in non-echo-chamber contexts. Make friends with people of different backgrounds, worldviews, life experiences. Listen to a variety of podcasts or watch a variety of videos. And so on.

Tip #11

Explore, don't expose. Use the tools of reason, universal experience, and genuine love/respect to *explore* the issues rather than *expose* flaws, cracks, or internal inconsistencies in your dialectical partner's beliefs.

Tip #12

Rid yourself of the 'game mindset'. There are no 'points' to be scored. There is no 'winning' (apart from gaining greater insight – which is *collaboration-based*, not competition-based).

Tip #13

Modify your language. The words we use influence the ways we think and feel. This is subtle but massively important. For example, use 'discussion', not 'debate'. Use 'here is something to consider', not 'here is why you're wrong'.

Tip #14

Take a virtue-based approach. Ask yourself: Am I cultivating the intellectual and moral virtues? Am I displaying the requisite intellectual humility? Your primary aim needs to be two-fold. First, aim for truth. Second, aim for loving each other. Virtue is all about aiming. Intellectual virtue is, fundamentally, aiming at truth. Moral virtue is, fundamentally, aiming at goodness.

Tip #15

Try not to sew your worldview or position into the very fabric of your own being. Your value is not tied to the positions you take. If you can (at least temporarily) separate yourself from the position you take, arguments against the position in question won't be seen or felt as personal attacks.

Tip #16

Listen. Be fully present. Put away distractions. But also put away your pre-conceived ideas about what the other person holds, believes, etc. In order to grasp their position correctly, you need to listen to *them*, not your pre-existing *conception* of them. This includes defining terms at the beginning of the discussion. Without defining terms, you only listen to your preconceived notions, not your dialectical partner.

If the occasion is right, re-state the point made by your dialectical partner. This shows that you are genuinely listening to them, which de-escalates tribalism. Also be sure to ask, respectfully, 'Is this what you mean?'

Tip #17

It's alright – in fact, *preferable* – to go slowly. Philosophy (and truth-seeking in general) is best done slowly. 'Clever', 'gotcha' type remarks only alienate your dialectical partner. It's better to avoid such 'gotcha phrases'.

Tip #18

This is a hard one: *steel-man*, don't straw-man.[4] You would be surprised at how much it reduces tension and tribalism within discussions when you take your dialectical partner's argument and improve upon it.

Tip #19

Use the principle of charity: when there are two ways to interpret something your dialectical partner said, focus on the one with greater plausibility or justification.

Tip #20

When they go low, go high. Don't stoop to levels of ridicule, insult, and tribalism. Maintain civility and respect. If it's appropriate, simply (respectfully) end the conversation or walk away.

Tip #21

Avoid psychologizing. Too often we spot potential psychological roots in *others* while failing to see the psychological roots of our *own* beliefs and worldview. Psychological influences on our beliefs are inescapable – it's part of the human condition. Focus instead on *reasons*, not psychological origins.

Tip #22

Arguments are not weapons and are rarely knock-down. Recognize room for rational disagreement. Use arguments as tools to probe the nature of reality. They're like illuminating flashlights. They don't force or compel agreement. They invite collective exploration.

4 Brace yourself…

So, that ends the first step on our journey. We've covered the importance of reducing tribalism, cultivating intellectual virtues, and engaging in productive dialogue. The next step on our journey will equip you with methods and tools for critical thinking and philosophical reasoning.

Chapter 2: Methods and Tools

1 Introduction

Before we can explore *how* to think critically, we need an understanding of *what* critical thinking is. Critical thinking is the process by which one utilizes the methods and tools of rational thought, evaluation, analysis, and inquiry in order to formulate reasoned conclusions and assessments.

Experts in critical thinking education define it in various ways. I'll give two extended quotations that capture it rather well.

> Critical thinking is the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action. In its exemplary form, it is based on universal intellectual values that transcend subject matter divisions: clarity, accuracy, precision, consistency, relevance, sound evidence, good reasons, depth, breadth, and fairness… (Scriven & Paul (1987))

Notice the emphasis placed on intellectual discipline – a core component of which is cultivating intellectual virtues. Elder (2007) emphasizes similar but complimentary elements of critical thinking.

> Critical thinking is self-guided, self-disciplined thinking which attempts to reason at the highest level of quality in a fair minded way. People who think critically attempt, with consistent and conscious effort, to live rationally, reasonably, and empathically. They are keenly aware of the inherently flawed nature of human thinking when left unchecked. They strive to diminish the power of their egocentric and sociocentric tendencies. They use the intellectual tools that critical thinking offers – concepts and

principles that enable them to analyze, assess, and improve thinking. They work diligently to develop the intellectual virtues of intellectual integrity, intellectual humility, intellectual civility, intellectual empathy, intellectual sense of justice and confidence in reason. They realize that no matter how skilled they are as thinkers, they can always improve their reasoning abilities and they will at times fall prey to mistakes in reasoning, human irrationality, prejudices, biases, distortions, uncritically accepted social rules and taboos, self-interest, and vested interest.

They embody the Socratic principle: The unexamined life is not worth living, because they realize that many unexamined lives together result in an uncritical, unjust, dangerous world (Elder (2007)).

With a better grasp of *what* critical thinking is, we are now in a position to explore how to do in the context of philosophical reasoning (and reasoning in general).

2 Methods and tools

2.1 Argument evaluation

Arguably the most important aspect of critical thinking is the ability to analyze and evaluate arguments. For this reason, the methods and tools that we'll explore will be understood within the context of argument analysis and evaluation.

A precondition of argument evaluation is argument analysis. To analyze something is to break it into its parts and to examine it in detail so as to determine the nature of and relations among its components. For argumentative purposes, this usually consists in formulating the argument in syllogistic form – meaning identifying the premises, the relations among them, and which conclusion (if any) follows from them. Be sure to break down the argument into its component propositions, and propositions into component terms or concept.

Following analysis is argument evaluation. While analysis serves mostly to give a better understanding of an argument, evaluation actually forms a *judgment* about its cogence and persuasiveness.

So, how does one evaluate an argument?

First, determine if the argument is *valid* or *invalid*. The latter means that the argument has a faulty structure. When a (deductive) argument is invalid, the truth of the premises does not guarantee the truth of the conclusion. A valid (deductive) argument, then, is such that the truth of the premises guarantees the truth of the conclusion.[5]

Here are some examples of invalid arguments:

Affirming the consequent

If A, then B.
B.
Therefore, A.

Denying the Antecedent

If A, then B.
~A.
Therefore, ~B.[6]

If the argument is valid, determine the truth value of its premises. If they are all true, then the argument is *sound*.

This dichotomy between validity and soundness is a valuable tool. Nevertheless, argument evaluation is much more complex than merely stating 'this argument is unsound.' We must specify the precise *ways* in which arguments can be unsound. And it's that very question that will be the focus of this chapter.

Each section in this chapter will examine ways to assess arguments. The sections will first describe the technique and then provide real-life examples

of how the technique can be applied.

Since there are so many ways an argument can go wrong, this chapter will mostly focus on *ways to critically assess arguments*. This chapter provides all rational thinkers with a rich (though not entirely exhaustive) tool kit for evaluating arguments.

2.2 Defeaters

You can either provide a *rebutting defeater* or an *undercutting defeater* for a premise, argument, or claim.

Roughly, a rebutting defeater shows that the premise is false (or likely false), thereby *refuting or rebutting* it. An undercutting defeater *removes or undercuts* the justification one has for accepting a premise. Undercutting defeaters can involve showing (i) the justification for a premise is unknowable, (ii) the justification does not adequately support the premise or conclusion, (iii) the premise is merely an unsupported assumption, or (iv) the justification for the premise relies on the truth of the conclusion of the argument (or the truth of the premise that is being questioned). An undercutting defeater essentially pulls the rug from under the argument, rendering it ineffective due to lack of proper substantiation. In general, if you want to undercut an argument, you can remove the justification for its premise(s), or you can remove the justification for its inference(s).

With these general types of defeaters in hand, we can now start to examine specific ways to evaluate arguments and claims. We will start with arguably the most important of all: showing the justification for an argument or premise to be *circular*.

2.3 Circularity

Circular motion was long thought to be the perfect form of motion, and as a result, it was postulated as the way heavenly bodies moved around the Earth. While circular motion may (or may not) be the 'perfect' form of motion, circularity is certainly far from perfect in argumentation. It is, in

fact, a pernicious fallacy that seeps into much of our reasoning and argumentation (many times unconsciously).

There are at least two types of circularity:

Circular Argumentation

In circular argumentation, one or more of the premises (or the justifications for the premises) presupposes the truth of the conclusion. It is essentially assuming the very thing in question. The conclusion is said to be 'included in' (either implicitly or explicitly) a premise, or a premise is said to be 'reliant upon' the conclusion.

This is also known as *begging the question* or *petitio principii*. It means assuming (whether explicitly or implicitly) from the get-go *the very question at issue* in the dialectical context at hand. A *dialectical context* is a fancy name for the conversational circumstances or situation in which you and your conversational partner disagree. Different dialectical contexts determine different standards of evidence and justification, and they also determine which assumptions, premises, and arguments count as question-begging and which don't. For instance, a court-room is a different dialectical context than a paper submitted to a mathematics journal. This is because the standards of evidence and justification differ in the two situations – the former allows circumstantial and testimonial evidence, whereas the latter restricts itself to strict mathematical proofs and demonstrations. Similarly, suppose you find yourself talking with a radical skeptic who disbelieves in an external world. In such a dialectical context, you cannot help yourself to assumptions which you would normally be allowed to help yourself – assumptions like 'my senses are generally reliable,' 'an external world exists,' and so on. Assuming these from the get-go would amount to merely presupposing the falsehood of your interlocutors position and hence begging the question against him or her.

Example

1. If humans are necessitated to act by prior conditions (i.e. if determinism is true), then giving each other praise and blame is unjustified.
2. Giving each other praise and blame is not unjustified.
3. Therefore, humans are not necessitated to act by prior conditions (i.e. determinism is false).

This is arguably an example of question-begging.[7] Premise (2) can only be justified by an appeal to free will, since the only way we can make sense of blaming person X for doing Y is if X could possibly have refrained from doing Y. By contrast, if X was necessitated by prior conditions to do Y, we clearly couldn't justifiably blame X (think of somebody sneezing – we cannot blame people for something they can't control). What this shows us, though, is that *whether* we are justified in blaming someone already *presupposes* we are justified in thinking that they are not necessitated by prior conditions. But that's *precisely the thing we set out to prove in the first place*. Hence, in the stipulated dialectical context, it seems we are in a position to accept (2) only if we already accept the conclusion.[8]

Circular Justification

If A justifies B, then you cannot also have your justification of B be A. Similarly, if A is justified by B, B is justified by C, but C is justified by A, then this chain of justification is *viciously circular*. Circular justification is essentially the same as circular argumentation, except circular justification refers to individual claims C whose justification rest on other claims C* whose justification, in turn, rest on C. It's a vicious circle: C is justified by C*, but C* is justified by C.

Example

'We know our senses are reliable because we all independently converge on the same beliefs about objects in the mind-independent world.'

The problem is, we could only justifiably claim that we all independently converge on the same beliefs about extramental objects if we *presuppose*

from the get-go that our senses are reliable. After all, if our senses are unreliable, then we clearly cannot reliably come to know – by means of our *senses* – that our fellow human beings agree with our claims about extramental objects. This proposed justification for the reliability of sensory perception therefore presupposes the very reliability of the thing needing to be demonstrated as reliable.

Example

Let's spice things up with a dialogue:

Payton: *The Da Vinci Code* is a true story.
Riley: How do you know?
Payton: Well, because it says it's true at the start.
Riley: But that is *assuming* that everything in it is true in order to *prove* that the book is a true story. You've appealed to the very thing in question – namely, the truth of what's written in the book – as proof of *itself*.

Example

Consider the following argument:

1. Philosophical zombies[9] are possible.
2. But if philosophical zombies are possible, then identity theory (a theory of mind which states that mental states are identical to physical brain states) is false.
3. So, identity theory is false.

Premise (2) is relatively straightforward: if it is possible for there to be a being that has every single molecule and atom in its brain entirely the same as yours but that lacks consciousness (mental states), your mental states cannot *be* your physical properties. In other words, mental states cannot be identical to brain states if it is possible to have one without the other. If A is B, then it cannot be possible to have A without also having B (and vice versa).

But – absent independent justification or demonstration – the first premise begs the question against the identity theorist.[10] To claim that it is possible for there to exist a philosophical zombie (i.e. a being whose brain is molecule-for-molecule identical to a human's but which lacks consciousness) *is just to claim* that neurophysiological states are not identical to mental states. In other words, *it is just to claim that the identity theory is false*. But that's the very conclusion that the zombie argument seeks to establish. One cannot argue from the possibility of zombies to the falsity of identity theory because the possibility of zombies *assumes* the falsity of identity theory.[11] This is because if identity theory is, in fact, true, then it is straightforwardly *impossible* for there to be philosophical zombies.

Be wary

At times, begging the question can be very difficult to spot. Consider the following passage:

> To allow every man an unbounded freedom of speech must always be, on the whole, advantageous to the State; for it is highly conducive to the interests of the Community, that each individual should enjoy a liberty perfectly unlimited, of expressing his sentiments (Whately (1826)).

In Whately's words, this passage is "abounding in synonymous expressions, which have no resemblance in sound, and no connection in etymology." His point is that each of the claims in this passage merely amount to *rephrasing* the same basic proposition; they say the same thing in different words. Appealing to one of them to provide reason or justification for believing the other amounts, then, to giving something as a reason for *itself*. In other words, it's circular reasoning.

The dependence of question-begging on dialectical contexts is captured rather nicely by the *Cambridge Dictionary of Philosophy*:

Whether or not the following argument about articles in this dictionary is circular depends on why the first premise should be accepted:

(1) The article on inference contains no split infinitives.
(2) [No] other articles contain… split infinitives.
Therefore, (3) No article contains split infinitives.

Consider two cases. Case I: Although (2) supports (1) inductively, both (1) and (2) have solid outside support independent of any prior acceptance of (3). This reasoning is not circular.

Case II: Someone who advances the argument accepts (1) or (2) or both, only because he believes (3). Such reasoning is circular, even though neither premise expresses just the same proposition as the conclusion. The question remains controversial whether, in explaining circularity, we should refer to the beliefs of individual reasoners or only to the surrounding circumstances. One purpose of reasoning is to increase the degree of reasonable confidence that one has in the truth of a conclusion. Presuming the truth of a conclusion in support of a premise thwarts this purpose, because the initial degree of reasonable confidence in the premise cannot then exceed the initial degree of reasonable confidence in the conclusion (*Cambridge Dictionary of Philosophy*).

Question-begging, then, is fraught with intricacies and complications. It's subtle – but with practice, you can learn to detect it with greater accuracy and precision. The *Internet Encyclopedia of Philosophy* echoes these intricacies and complications:

Some logicians suggest that, in informal reasoning with a deductively valid argument, if the conclusion is psychologically new insofar as the premises are concerned, then the argument isn't an example of the fallacy. Other logicians suggest that we need to look instead to surrounding circumstances, not to the psychology of the reasoner, in order to assess the quality of the

argument. For example, we need to look to the reasons that the reasoner used to accept the premises. Was the premise justified on the basis of accepting the conclusion? A third group of logicians say that, in deciding whether the fallacy is present, more evidence is needed. We must determine whether any premise that is key to deducing the conclusion is adopted rather blindly or instead is a reasonable assumption made by someone accepting their burden of proof. The premise would here be termed reasonable if the arguer could defend it independently of accepting the conclusion that is at issue (Dowden).

For our purposes, we will treat question-begging or circularity as accepting a conclusion for reasons that implicitly or explicitly rely on that *very conclusion itself*. The best ways to grasp circularity, though, are (i) to understand paradigm cases of it, and (ii) to continually practice spotting it in lines of reasoning.

2.4 Strategies

Like the section on tips for productive conversations in the previous chapter, this section will consist in a numbered list of strategies for critically evaluating arguments. Some strategies will be fleshed out in great detail while others will be brief. The purpose, again, is to equip you for thinking critically in philosophical contexts.[12] We will apply these strategies extensively in later chapters when we analyze different issues in philosophy of science and philosophy of mind.

Strategy #1

Show that a premise, argument, or assumption conflicts with well-established empirical facts.

Strategy #2

Show that a premise, argument, or assumption entails something contradictory, false, absurd, or something conflicting with the arguer's other beliefs and/or with the very argument in question. Perhaps the premises,

when applied to some other aspect of the world, imply something that would undermine or rebut the arguer's position.

Strategy #3

Show that a premise or assumption is unjustified/unwarranted, or demonstrate that any underlying assumptions are implausible, underdeveloped, unsubstantiated, or controversial.

Example

'You haven't even died yet, so you cannot know what heaven or hell are like or if they even exist.'

This contains an underlying assumption that direct experience is the only means by which one could gain knowledge of this sort, which is a problematic and unjustified assumption. Some would argue that if theism is true, it could easily be the case that some miraculous event led God, an angel, or a saint to reveal (some of) the conditions of heaven and/or hell. Without *justifying* this underlying assumption, the argument in question does not succeed.

Strategy #4

The thing being argued for (such as a criterion for reliability, knowability, or truth) refutes itself. It is said to be self-refuting, self-defeating, or self-undermining.

Example

'You ask me to disprove the existence of fairies, but that's impossible – you cannot prove a negative.'

This argument, though, is self-defeating. For, if one could prove[13] that you *cannot* prove a negative, one would *thereby have proven a negative*. One would have proven that *it is not the case that* a negative can be proven.

Thus, if one could prove that very statement, one would have demonstrated its falsity. Thus, it is self-defeating.

More fundamentally, though, you *can* prove a negative for independent reasons. There is an entire law of logic dedicated to proving negatives (the law of non-contradiction), which states that it is not the case that anything can be both A and ~A in the same respect at the same time. Thus, we can easily prove that it is not the case that, say, there exists a square circle. For that would amount to something being a square and not a square, which is impossible. So, it is false to suppose that a negative cannot be proven.[14]

Strategy #5

Show that one or more premises has a justification that is undercut by some further consideration. This means that we should not accept the premise on the basis of the provided justification.

Example

Some argue there are instances of intense suffering that an omnipotent, omniscient, wholly good being could prevent without thereby losing some greater good or preventing some evil equally bad or worse. One justification commonly leveled in support of this claim is as follows: find a token or type[15] of evil state of affairs and then argue that, because we cannot conceive of a morally justifying reason for this evil to occur, there *is* no such reason (else: there is *probably* no such reason). One would then conclude that such evil is *gratuitous* (pointless, unnecessary, unwarranted, unjustified).

Skeptical theism provides an example of the type of undercutting defeater described in strategy #5. Skeptical theism aims to undercut the inference from 'I do not see or conceive of any morally justifying reason' to 'therefore, there (probably) is no such reason'. It points out that this 'noseeum inference' can only be justified under certain conditions.

Take, for instance, the example of a doctor dropping a needle on the floor, picking up the needle, and proclaiming, 'I do not see any germs on the

needle, and hence there (probably) are no germs.' This is a poor noseeum inference, because *if there were* germs, then we would not expect to see them. In fact, if there *were* germs present, we would positively expect *not* to see them!

Another example of a noseeum inference is as follows. If you walk into your living room and see no elephant, you can justifiably say 'I see no elephant in the room, therefore (probably) there is no elephant in the room'. However, you cannot justifiably reason as follows: 'I see no dust in the room, therefore (probably) there is no dust in the room'. This is because there are certain conditions that must obtain for adequate epistemic access to certain truths, and we must meet such conditions of reasonable epistemic access in order to reliably and justifiably make a noseeum inference. Due to the nature of dust, we do not meet the conditions of epistemic access that would allow us to justifiably make this noseeum inference.

The inference in the examples above is of the form:

'I do not (else: cannot) see/conceive of X, therefore (probably) there is no X.'

However, we need to ask: do we have adequate epistemic access to X in order to reliably and justifiably make this inference? If there *were* an X, would we expect to see it? If there *were* germs on the needle, would we expect to see them? If there *were* dust in the room, would we expect to see it? If there *were* a morally justifying reason for God allowing certain evil (tokens or types of) states of affairs to obtain, would we expect to see such a God-justifying reason?

Skeptical theists (in short) argue that because there is such a staggering and vast cognitive chasm between our finite intellects and God's infinite intellect, we *shouldn't expect* to see/know his justifying reasons for everything he does/allows. We do not have representative epistemic access to God's morally justifying reasons, and thus the inference sketched at the beginning of this example is a poor one. Skeptical theists will also emphasize that this cognitive chasm means that we humans do not have

representative knowledge of the full range of possible goods and evils and the logical and causal connections among them – and that, because of this, we cannot justifiably arbitrate on whether or not there are morally justifying reasons present.

However, this is just a single (but very illustrative) example of a more general way to undercut an argument. Whether or not skeptical theism succeeds is beside the point. The point is, rather, to illustrate a real example of undercutting defeaters in action.

Strategy #6

Show that definitions, terms, or propositions within the argument are faulty, vague, or ambiguous.

Example

'Science and religion are incompatible.'

What does this mean by science? Does it mean the *process* of science, its methodology? Its assumptions? Its ways of knowing and thinking? The body of evidence that science has uncovered? Scientific theories? Which ones?

What does this mean by religion? Does it mean religion as a sense of spirituality and a mere belief in a transcendent being? Or does it mean a particular religion that makes certain empirical claims? Which ones?

What does it mean by incompatible? Does it mean contradictory? That the two *probably* cannot co-exist? Or that religious claims have to be molded to fit the science rather than being outright concordant with them?

Without properly defining terms, this claim is underdeveloped.

Example

'Flying planes can be dangerous.'

Is the *act of piloting and flying a plane* dangerous? Or are the *planes themselves that are flying* the ones that are dangerous?

Strategy #7

It would be entirely remiss of me to ignore the various informal fallacies that are necessary for evaluating arguments. Another way to evaluate arguments, then, is to show that the argument (or premise) commits any number of informal fallacies, such as false equivalence, false dichotomy, special pleading, composition fallacy, and genetic fallacy.

Example

Let's take the fallacy of division as an example. It involves taking an attribute of a whole (or class) and assuming that it must necessarily apply to each part (or member). Structurally, the fallacy is as follows: A has property X. So, all parts (or members) of A have this property X.

Here are some fallacious instances of this pattern: Zootopia is the richest country in the world. So, everyone in Zootopia must be rich. The judicial system in Zootopia is fair. So, the defendant received a fair trial.

Curiously, though, there are valid arguments with a similar pattern: Human beings are mortal, and so Socrates is mortal. Why are some instances of the pattern fallacious while others aren't?

The difference lies in the distinction between *distributive* and *collective* attributes. Attributes that are shared by all members *by virtue of being a member* are called distributive. Attributes that are created only by bringing together the right parts (members) in the right way are *collective*. The difference can be illustrated by the difference between:

(1) Planets are large.
(2) Planets are numerous.

In (1), the attribute *large* is distributive. It is a quality held by each planet individually in virtue of the *kind* of thing it is. In the second sentence, the attribute *numerous* is collective. It is an attribute of the *group* of planets that exists only because of the *collection*. No planet can individually be 'numerous'. When things are brought together to compose a whole, they can often result in something with properties not possessed by the parts individually.

Example

'Because each cell in your brain is incapable of consciousness, the consciousness in your brain cannot be explained by matter alone.'

But, absent an independent reason to think that consciousness is something that must be possessed by the parts if it is possessed by the whole, this commits the fallacy of division.

Example

'Evolutionarily-instilled dispositions explain why religion exists. It is merely adaptively advantageous and serves the purpose of social cohesion. Therefore, since it is merely adaptive and not truth-seeking, one ought to reject it.'

This is a genetic fallacy. Merely from the fact that religious belief originated from evolution or adaptive behavior, it doesn't follow that religious belief is thereby false. The origin of something does not thereby make it true or false – it's truth or falsity is a function of whether or not it corresponds to the way the world is in reality.

What's more, this type of argument is self-defeating. Consider that (on this account) *all our cognitive faculties* originated by means of evolution. Hence, if *X originating by means of evolution* is sufficient for showing X to be false or unreliable, it follows that *none* of our cognitive faculties are reliable. But, of course, this is self-defeating, since in order to justify this very claim one must presuppose the reliability of his or her cognitive faculties.

Strategy #8

If an argument by analogy is offered, show that there are clearly and relevantly disanalogous elements between the analogy (on the one hand) and the thing that the analogy is being used to support (on the other).

When using analogies, one must justify how very specific relationships within two different concepts or situations are related and/or function similarly to one another, as well as justify that those relationships actually exist within each concept or situation.

Assessing arguments from analogy, then, requires determining whether there are *relevant dissimilarities* between the two analogues that render the desired inference faulty.

Things can be *relevantly analogous* (in which case the analogy does, in fact, support the point being made), *relevantly disanalogous* (in which case the analogy does not, in fact, support the point being made). Things can also be *irrelevantly analogous* and *irrelevantly disanalogous*.

If someone attempts to pinpoint a disanalogous element within an analogy, but yet that disanalogous element is *irrelevant* to the point being made, the force of the original argument from analogy remains. In this case, the objector points to an *irrelevantly disanalogous* element of the situation.

Likewise, the similarity between the two things in question may be *irrelevant* to the desired conclusion and hence does not actually support the point being made. In this case, the respect in which it is analogous is *irrelevant*.

Strategy #9

Show that a criticism of an argument merely 'kicks the can back a step'. This occurs when someone aims to rebut or undermine an argument by appealing to some other, further fact F, but in reality, this just kicks the can back a step and the argument still applies to F. In such a scenario, the original argument retains its force.

Example

A number of authors have argued that the hundreds of millions of years of animal pain and suffering in the evolutionary process pose a problem for belief in God's existence.

In response, one might argue that 'All animals except for primates lack higher order consciousness, that is to say, an awareness of their own pain. Therefore, the problem of evolutionary animal suffering is not a problem.'

Notwithstanding this claim's deep implausibility, there have *still* been primates for tens of millions of years. It follows, then, that this response merely pushes the problem back a step: for the argument will simply be based on the tens of millions of years of *primate* suffering. Postulating this 'primate suffering hypothesis' merely kicks the can back in an attempt to sidestep the argument, but the argument resurfaces.

Strategy #10

Suppose Jones argues that we would expect some evidence or observation O_1 given some hypothesis H. Suppose, further, that O_1 obtains, and that Jones argues (on this basis) that H is probably true. One way to assess arguments along these lines is to point out that, under H, we would *also* expect to see O_2 and that O_2 does *not* obtain (thereby providing evidence against H). Additionally, you could point out that, under H, we would *not* expect to see O_3 obtain and that O_3 *does* obtain.[16]

Example

Let's remain on the topic of the problem of animal suffering. Some respond to the argument by pointing out pain's crucial biological role in causing animals to act in ways that are self-preserving. As a result, pain is necessary and justified (so the rejoinder goes).

Of course, one problem here is that this response merely *presupposes* the existence not only of natural disasters and things which cause suffering but

also of threatening organisms that voraciously predate on other organisms.
[17] But the problem of animal pain is, at least in part, captured by the
following question: why are there such evil things that cause suffering *in
the first place*?

But, more importantly (and more relevant to strategy #10), if the morally
justifying reason for creating pain is its role in guiding biological organisms
to self-preserving ends, there seem to be problems along the following
lines: (i) there are observations we would *expect* under such a hypothesis
that do *not* obtain, and (ii) there are observations we would *not* expect
under such a hypothesis that *do* obtain.

First, some instances of suffering in human and non-human organisms (and
some diseases for that matter) cause pain that doesn't affect behavior in any
clear ways. In such cases, the organisms suffer without being guided
towards biologically self-preserving ends. This is an example of pain which
doesn't serve a biological end – something that is unexpected on the
hypothesis on offer.

Second, some instances of intense suffering occur as an organism dies, such
as a fawn being burned in a forest fire, or an elephant being eaten alive by
lions. Such pain clearly serves no self-preserving end, since there is no self
to be preserved in the process of dying an agonizing death.

Third, under the proffered hypothesis, we would *expect* to see certain
observations that we do not, in fact, see. For example, some instances of
biological ends which *would* be crucial to self-preservation are associated
with little to no pain – breathing in carbon monoxide, for instance.

Hopefully this example gives you a sense of strategy #10's significance and
applicability.[18]

Strategy #11

Statistical and empirical arguments are crucial to the formulation of
justified beliefs. As a result, knowing how to evaluate such arguments is of
paramount significance.

One way to assess statistical arguments is to determine whether they contain hasty generalizations (from a small sample population, say, or a single experiment or observation). Also determine whether the statistic in question accounts for confounding variables.

To think critically about such experimentation and research, ask the following questions:

Was the study conducted in the laboratory or in a natural setting? Was the sample representative across time, culture, socioeconomic status, mood, age, sex, etc? Were all possible confounding factors accounted for?

In order to critically analyze psychological and statistical findings from experiments, do the following:

- Identify the aim

Identify in the procedure:

- Participants: sampling method, sample size, and variables in the experiment
- Type: Questionnaire? Survey? Observational? Experimental?
- Method: Was the experiment a *repeated measures* experiment (i.e. the same sample group with different conditions applied)? Was it a longitudinal study (over a long period of time)? Was it an *independent measures* experiment (i.e. one group tested with one condition and another group with another condition)? Was the study cross-cultural?

Identify findings:

Actual results, numerical results (exact numbers, trends, or correlations)

Identify conclusions:

- Generalization(s) drawn from data
- Evaluate for: internal validity; benefits and drawbacks of sample; controls; demand characteristics (demand characteristics occur when subjects consciously or subconsciously respond the way (the subjects believe) the researcher's desire – they also occur when subjects realize what factor is being observed and change their behavior accordingly); and social desirability effect (the subjects know they're in a study so they desire to appear better or more desirable than they in fact are)
- Extraneous/confounding variables (other things that may have influenced results)
- External validity: how results can be generalized outside the context of the particular experiment
- Ecological validity (behavior across the population) is lower in a lab than in the real world
- Locational validity (based on the particular sample and population being observed)
  - A study is more locationally valid if it tests across a variety of backgrounds and cultures
- Temporal validity: cultures, individuals, and trends (psychological or otherwise) change over time
- Cultural dimension: different perspectives and values within cultures can introduce confounding factors into the experiment

Findings

- Were the participants asked to do things that are far from real life?
- Are the findings of the study supported/questioned by the findings of other studies?
- Do the findings have practical relevance? To what extent? What are the limitations?
- Variables in an experiment need to be operationalized – that is, they need to be written in such a way that it is clear *what* is being measured. Be specific! Don't just say 'noise'. Say the operationalized variable of high music at volume 35. Don't just

say 'recall'. Say the operationalized variable of number of words remembered from a list of 20 words after 15 minutes.

Sampling

- What is the nature of the sample?
- Is it an opportunity (convenience) sample? This is a sample of whoever happens to be around and agrees to participate – for example setting up your station in a supermarket. But we must ask: What type of people shop at the supermarket? What type of people say yes to participating? Is there a gender imbalance in who does the shopping in this community?
- Is the sample self-selected? If so, why suppose that those who volunteer properly represent the population at large?
- It is important, also, to consider participant variability – that is, the extent to which the participants may share a common set of traits that can bias the outcome of the study. If a survey is about anxiety about mathematics, it may be the case that, if you do a volunteer study, only people who feel very strongly and have a lot of anxiety about math may volunteer, thus skewing results.
- Generally, the best sampling method is random sampling. A random sample is one in which every member of the target population has an equal chance of being selected.

Variables

Confounding variables are variables that influence the relationship between the independent and dependent variables of an experiment.

- *Demand characteristics* occurs when participants act differently simply because they know they are in an experiment.
- *Researcher bias* occurs when the experimenter (selectively) sees what he or she is looking for. The expectations of the researcher may consciously or subconsciously affect the findings of the study. Use a double blind study to help combat this. In this

design, not only do the participants not know whether they are in the treatment or control group, but the person administering the treatments and placebos knows neither the aim of the study nor which group is the treatment/control.

- *Participant variability* occurs when characteristics of the sample affect the dependent variable.

Bidirectional ambiguity is present when two variables are correlated but whether the first causes the second or the second causes the first is ambiguous, indeterminable, or unknown. In a mere correlational study, since no independent variable is manipulated it is (nearly) impossible to know if x causes y, y causes x, if they interact to cause behavior, or whether the relationship is just coincidental.

Be cognizant, moreover, of *researcher degrees of freedom*. This "means choosing when to stop [and start] recording data, which variables to follow, which comparisons to make, and which statistical methods to use" (Novella (2018)). Now, all of these are key components of nearly every scientific study. If, however, the researchers monitor the data or outcomes in any way *while* making these meta-observational choices, they can consciously or subconsciously exploit their degrees of freedom to reach *apparently* statistically significant results. Anything a researcher changes about the research during or after collecting the data introduces bias.

There are a number of reasons why we explored the critical thinking basics concerning statistical and experimental reasoning. First, a number of substantial philosophical theses are argued for (in large part) on the basis of empirical experimentation.[19] Second, they make you a good consumer not only of products but also of information more generally.

Strategy #12

Show that a proposed explanation either explains nothing, rephrases (or presupposes) the very thing in need of explanation, or is *itself* in need of

explanation.

Example

'Analgesics work because they reduce pain.'

The thing is, an analgesic is defined as a chemical substance that *reduces pain*. So, its pain-relieving property cannot be explained by saying it 'relieves/reduces pain'. That explains nothing.

Strategy #13

Show that an argument, premise, or inference conflates epistemology with metaphysics.[20] This is also known as the 'masked man' fallacy.

Imagine you're at a masquerade ball and everyone is wearing masks. There is a man that catches your attention, however he has been wearing a mask the entire time.

Your friend tells you, 'Hey, I think that is your father!'

Unconvinced, you retort, 'Well, that's false. I know who my father is, but I don't know who that masked man is. Therefore, their properties are different, and therefore they cannot be the same person.'

This is fallacious. What you *think* about something or someone is a property of *you*, not a property of the person or thing *itself*. Conflating the two is the masked man fallacy.

Example

1. I can doubt that Hesperus (evening star) is Phosphorus (evening star)
2. I cannot doubt that Phosphorus is Phosphorus.
3. Therefore, Hesperus is not Phosphorus

The conclusion is said to follow from premises (1) and (2) because the Hesperus and Phosphorus have different properties – namely, one has the property *being dubitable by me* while the other doesn't. And if two things are identical, they cannot have different properties. This is an appeal to Leibniz's law, which states that if A and B are the same thing (numerically identical), then any property of A must also be a property of B (and vice versa).

But on the supposition that this is a good argument, we get the absurd result that Hesperus is not Phosphorus.[21] One way some have accounted for the argument's failure is by arguing that properties like *being dubitable by me* are not genuine properties of things in themselves. Another way is to distinguish between *intensional* and *extensional* contexts. But don't worry – we'll cover this near the end of the chapter.

Strategy #14

Show that the argument in question succumbs to a 'parody argument' that exactly parallels the structure and inference patterns contained in the original argument but which either (i) contradicts the conclusion of the original argument, (ii) establishes the truth of something known to be false or something taken to be absurd, or (iii) entails the falsehood or unjustified nature of one or more premises in the original argument. Parody arguments are central in philosophy, and they force us to consider *symmetry breakers* – relevant differences or considerations that make only *one* of the parallel arguments justified. They break the symmetry (as it were) between the seemingly exactly parallel (i.e. symmetrical) opposing arguments or considerations.

Example

One version of the modal ontological argument for the existence of God runs as follows:

1. Possibly[22], God (a necessarily existent, perfect being) exists.
2. So, God – necessary being – exists in some possible world.[23]

3. But a necessary being could exist in *any* world only if it exists in *all possible worlds* – that's the definition of a necessary being (a being that cannot fail to exist, i.e. exists in all possible worlds).
4. So, God exists in all possible worlds.
5. But the actual world is a possible world.
6. So, God exists in the actual world.
7. If God exists in the actual world, God exists.
8. So, God exists.

Or, more simply:

1. It is possible that God exists.
2. If it is possible that God exists, then it is necessary that God exists.[24]
3. So, it is necessary that God exists.

While these arguments are valid[25], they succumb to (at least) one major difficulty: a symmetrical, parallel argument. Consider the following arguments:

1. Possibly[26], God (a necessarily existent, perfect being) doesn't exist.
2. So, God – necessary being – doesn't exist in some possible world. [27]
3. But a necessary being could exist in *any* world only if it exists in *all possible worlds* – that's the definition of a necessary being.
4. So, God couldn't exist in *any* world.
5. But the actual world is a possible world.
6. So, God couldn't exist in the actual world.
7. If God couldn't exist in the actual world, God doesn't exist.
8. So, God doesn't exist.

Or, more simply:

1. It is possible that God doesn't exist.
2. If it is possible that God exists, then it is necessary that God exists (i.e. it is not possible that God doesn't exist).[28]
3. So, it is not possible that God exists.

Like the first two arguments, the third and fourth arguments are valid. Strangely, though, the former two arguments have a conclusion which *contradicts* that of the latter two arguments. It *cannot* be the case, then, that both arguments are sound.

The crucial premise in each argument is the *first premise* – the 'possibility premise'. But because the 'possibility premises' in the contradictory arguments (i) exactly parallel one another, but (ii) cannot both be true, it follows that *if* one wants to use one of the premises as an argument for (or against) God's existence, one needs to offer a *symmetry breaker* – some principled, non-arbitrary, non-question-begging reason to favor one possibility premise as opposed to the other. Absent such a relevant difference or symmetry breaker, preferring one possibility premise over the other amounts to an unjustified metaphysical prejudice.

Strategy #15

Let's cover two strategies at once.

First, give a counter-example to the claim or analysis. A counter-example is some fact, event, proposition, or state of affairs that runs contrary to a premise, conclusion, hypothesis, worldview, or proposition. For instance, if someone says X is identical to, defined as, or analyzed/understood in terms of Y, see if you can have Y without X or X without Y.

Second, show that a claim or inference is a *non-sequitur* (i.e. does not follow from the given information or propositions).

Example

The following is not an argument against eating meat; rather, it's a string of arguments that serves to illustrate the *counter-example* and *non-sequitur*

strategies. The arguments presented are by no means representative of all the arguments on either side of the debate.

Suppose Jones reasons: 'Animals taste good, so it's morally permissible to eat them.'

- If humans tasted good, would it be morally acceptable to kill and eat them?[29]
- Under this moral justification, it would seem to follow that anything that tastes good is morally acceptable to kill and eat.
- Moreover, this argument may conflate *causes* with *good reasons* (i.e. justifications). Tasting good may be the causal explanation for why you eat meat, but it does not *morally justify* that practice.

Jones: 'Well, it's natural to kill animals for food, so it's morally permissible to eat them.'

- This is an appeal to nature fallacy. Even if eating meat were natural, it does not follow that it is good or morally permissible. [30]
- Poisonous berries are in some sense natural, but they are not good in the sense of facilitating our health and flourishing. Person X may have 'naturally' violent tendencies or be naturally self-centered, but it doesn't follow that it is good or morally permissible for Person X to act on these tendencies.
- We cannot infer goodness or badness from the mere fact that X is natural or unnatural. An additional linking premise is required for such an inference to be valid.
- There are acts that seem both natural and immoral. Many biologists argue that evolution has instilled within humans (or a proper sub-set of humans) predispositions towards aggression and xenophobia.

- There seem to exist instances of all of the following combinations:
  - Natural + Good (empathy, water, love)
  - Natural + Bad (aggressive tendencies, predation, lying, poisonous berries)
  - Unnatural + Good (vaccines, modern medicine, clothes, housing)
  - Unnatural + Bad (lethal injections, CFC destruction of ozone layer)

Jones: 'Humans are on top of the food chain, so it is morally permissible to kill and eat animals lower on the food chain.'

- This is a subtle form of the appeal to nature fallacy, as it assumes that the chain that exists in natural reality is necessarily connected to moral permissibility.
- This argument is based on the principle that *might makes right*. However, this principle seems false. Consider Person Y. Unfortunately, Y mistreats animals. Merely from being more powerful or mightier, or merely from the fact that Y is higher on the food chain than these animals, it does not follow that Y's mistreatment is morally permissible. Merely from the fact that someone is mightier or stronger or physically superior to another, it doesn't follow that his or her actions are morally justified.
- If some animal evolved to be higher than humans, would it be morally permissible for them to slaughter and consume humans? If an advanced alien race visited earth, and if they were more powerful, intelligent, rational, and advanced, would they be morally justified in capturing us, killing us, farming us, and eating us?

Jones: 'We need meat for nutrients and protein, and thus it is morally permissible to kill and eat animals.'

- First, meat is not essential to a healthy diet and lifestyle, so it is not a necessary condition for health.
- *Even if* we need meat, that does not by itself justify killing, farming, and consuming animals like pigs and cows as opposed to lower animals like fish and insects (or laboratory-grown meat, something likely to revolutionize the future).
- Even if you need meat, it does not necessarily justify the *kind* and *quantity* of meat we eat, or the way in which the animals are *treated/farmed/killed*.

Jones: 'Consider a situation in which you are trapped on a lifeboat. You could throw a cow off the boat, or do the same to an intellectually challenged human being. Which would you choose?'[31]

- This argument aims to show that, intuitively, human life is more valuable than animal life. But from the mere fact that human life is more valuable, it doesn't follow that animals utterly[32] lack moral value. So, even if a cow has *less moral value* than a human, it still needs to be justified as to why their moral value is such that killing and eating them is morally justified.
- Most of us value our family members more than strangers. Would that justify killing and eating strangers? From the mere fact of valuing X more than Y, it doesn't follow that we can treat Y radically different than we treat X. Pinpointing a morally relevant difference between X and Y may allow one to justifiably treat them differently in *some ways*, but a justification must be presented regarding the *extent* of different treatment.
- How we (ought to) act in an extreme case when survival is on the line is not necessarily a reflection of what is morally permissible when such extreme conditions are not met. Most first-world people, for instance, are not in such extreme situations. So, *even if* it is acceptable to kill and eat a cow for survival, it does not follow that it is always and in all cases morally permissible to kill and eat a cow, especially when dire survival is not on the line.

- Finally, this argument is an instance of the absolutist fallacy. Absolutism arises when we make no exceptions for rules that have (or ought to have) exceptions. For instance, it is usually wrong to lie, but it is morally permissible to lie to the Nazis about the Jews in your basement. Similarly, one might hold that it is usually wrong to kill and eat animals, but not in all cases.

Jones[33]: 'Animals are unintelligent and non-rational, and it is thus morally permissible to kill and eat them.'

- Many animals exhibit suffering, forms of metacognition, and problem-solving.
- Even if they are less intelligent or less aware in some way, it does not follow that it is morally permissible to kill and eat them. Is it morally permissible to eat severely mentally handicapped people? Infants? Fetuses? Alzheimer's patients? What if one farmed humans and genetically engineered the fetuses to have the intelligence at or lower than a pig?

For present purposes, the *content* of the arguments is less important than the *strategies* of finding counter-examples and pinpointing non-sequiturs.

Strategy #16

Explanatory accounts of certain features of reality must fit specific criteria to count as genuine explanations. An explanation E of fact F cannot itself presuppose the very thing to be explained (F).

Furthermore, an explanation E of fact F, if it is to be a genuine explanation, cannot entail a vicious infinite explanatory regress or a vicious explanatory circle. Suppose I want to explain why there are three eskimos (E_1, E_2, and E_3) at the bus stop. Suppose we have an explanation of why all three of them are there: E_1 is at the bus stop because E_2 is there (perhaps he followed him or walked beside him), E_2 is there because E_3 is there, and E_3 is there because E_1 is there. In this scenario, we end up *not explaining why*

they are at the bus stop in the first place. We have merely explained the *relations between* the eskimos and thus have not given any account as to why they all are at the *bus stop* rather than (say) the *supermarket* (or for that matter why they even moved from their original starting point at all). This is a viciously circular explanation.

So, three common fallacies that explanations need to avoid are: (i) assuming the existence of (or making use of) the very thing needing to be explained, (ii) engendering a viciously circular explanation, and (iii) engendering a vicious infinite regress of explanation.

Example

In this example, philosopher Edward Feser examines the *conceptual role account* of intentionality.[34] He writes:

> This sort of theory proposes that the meaning or intentional content of any particular mental state (a belief, desire, or whatever) derives from the role it plays within a system of mental states, all of which… seem logically interrelated, since to have any one mental state seems to require having a number of others along with it.
>
> If we think of beliefs, desires, and the like as a vast system of logically interconnected elements, the theory holds that each element in the system gets its meaning from having precisely the place in the system it has, by bearing exactly the logical and conceptual relations it bears to the other elements.
>
> There seems to be a serious problem with the conceptual role approach, namely that even if it is granted that mental states have the specific meaning or content they do only because of their relations to other mental states, this wouldn't explain how mental states have any meaning at all in the first place. That a particular belief either implies other beliefs or is implied by them presupposes that it has some meaning or other: nothing that was

completely meaningless could imply (or be implied by) anything. The very having of logical and conceptual relations assumes the prior existence of meaning, so that no appeal to logical and conceptual connections can (fully) account for meaning.

Moreover, if belief A gets its content from its relations to beliefs B and C, and these get their content from their relations to beliefs D, E, and F, we seem destined to be led either in a circle or to an infinite regress. Either way, no ultimate explanation of intentional content will have been given. To provide such an explanation thus inevitably requires an appeal to something outside the network, something which can impart meaning to the whole (Feser (2007)).

Don't get too bogged down in the philosophy of mind – just try to appreciate the strategies Feser employs for critically engaging with the conceptual role theory of intentionality.

Strategy #17

Be very careful with claims about conceivability and imaginability.

Consider the following claim: 'One can imagine of a bowling ball popping into existence without a cause; hence, such a scenario is metaphysically possible.'

But there is a subtle difference between *not being able to imagine X* and *positively being able to imagine X not to be the case*. Another distinction to be made is between *conceiving/imagining X without also conceiving/imagining Y* and *conceiving/imagining X without Y*.

To illustrate this last point, take the example of the bowling ball. There is a crucial difference between imagining a bowling ball coming into existence *without also imagining its cause* and actually imagining the bowling ball coming into existence *with no cause*.

The latter is, in fact, incredibly difficult to do while the former is simple. To imagine a bowling ball actually coming into existence with no cause

whatsoever, you would have to distinguish between such an imagining (on the one hand) from (on the other hand) imagining a situation in which a bowling ball is merely *transported* from another place (seemingly out of thin air).

Details aside, the point is that imagining X *without also* imagining Y is *not identical* to positively imagining X existing without Y.

Strategy #18

Show that the argument or inference pattern commits a *scope fallacy* (also known as a quantifier shift fallacy).

Consider the claim *every girl loves a puppy*.[35] This is ambiguous between:

> (1) For every girl, there is a puppy (perhaps a different puppy for each girl) whom that girl loves.
> (2) There is a certain, individual puppy whom every girl loves.

In logical notation, the distinction can be captured as follows:

> (1) (For each girl x)(there exists a puppy y such that)(x loves y)
> (2) (There exists a puppy y such that)(for each girl x)(x loves y)

In (1), the expression 'for each' has *wide scope* while the expression 'there exists' has *narrow scope*. In (2), this order is reversed.

The same distinction arises with modal quantifiers (like 'possibly' and 'necessarily'). For instance, consider the claim that *the inventor of the wheel necessarily invented the wheel*. Once again, this is ambiguous between:

> (1) It is necessary that the inventor of the wheel invented the wheel.
> (2) The specific inventor of the wheel necessarily invented the wheel.

In logical form:

(1) Necessarily, if x invented the wheel, then x invented the wheel.
(2) If x invented the wheel, then it is necessary that x invented the wheel.

In (1), the necessity operator has wide scope, whereas in (2), the description 'x invented the wheel' has wide scope.

And note that these two mean wildly different things. Statement (1) is obviously true. It is of the form 'necessarily, if P, then P'. But statement (2) seems blatantly false – the actual person who invented the wheel did not *have* to invent the wheel – such a person could have died young, or pursued his or her passion to become the inventor of the axle instead of the wheel, or what have you.

The fallacy arises when one subtly infers (2) from (1) in the cases described above. Merely from the fact that every girl is such that she loves a puppy, for instance, it does not follow that there is some single puppy that all girls love. That's like inferring from *every human is such that he or she has a mother* to *there is some individual mother for all humans*.

Example

Examples of quantifier shifts are subtle but crucially important, especially in philosophy of religion. Below are three examples of quantifier shifts.

First, suppose Jones mounts the following argument:

1. Some things change (i.e. there are changes).
2. Whatever changes is caused to change by something actual.
3. Therefore, some things are caused to change by something actual.
4. If some things are caused to change by something actual, then there are chains of changes (i.e. one thing changed by another, in turn changed by another, and so on).
5. Therefore, there are chains of changes.
6. But chains of changes cannot be infinitely long.

7. If chains of changes cannot be infinitely long, then such chains terminate in one first member (the unchanged changer, unmoved mover, or unactualized actualizer).
8. Therefore, such chains terminate in one first member (the unchanged changer, unmoved mover, or unactualized actualizer).

With quantifier shifts in mind, though, consider premise (7): If chains of changes cannot be infinitely long, then such chains terminate in one first member. Notice that there seems to be an ambiguity here. Jones is either arguing (1) because each individual chain must terminate in a first member, each individual chain has one first member that serves as this chain's termination; or (2) because each individual chain must terminate in a first member, there is one first member into which all chains terminate Look at the structure of these closely:

(1) Each chain is such that there is one first member for that chain.
(2) There is one first member for all chains.

This clearly resembles the (1)'s and (2)'s in our explanation of strategy #18. Importantly, all Jones has demonstrated[36] is (1) – namely, that since chains cannot be infinite, any individual chain must have (at least) one first member. But just as any individual student's having a counselor does not entail there is a *single counselor* that all the students have, the fact that each individual chain has its own termination does not entail that there is *one termination* to which each and every chain traces back. Claim (2) simply does not follow from claim (1).

Second, consider the following line of reasoning from Thomas Aquinas:

> We find in nature things that are possible to be and not to be, since they are found to be generated, and to corrupt, and consequently, they are possible to be and not to be. But it is impossible for these always to exist, for that which is possible not to be at some time is not. Therefore, if everything is possible not to be, then at one time there could have been nothing in existence (ST I, q2, a3).

Here, Aquinas argues that (1) entails (2) (see below), thereby committing a quantifier shift fallacy. Let 'thing which is possible to be and not to be' be called a *corruptible thing*. Here:

(1) Each corruptible thing is such that it doesn't exist at some time.
(2) There is some time such that each corruptible thing doesn't exist at that time.

Once again, (2) simply does not follow from (1). It could be the case, for instance, that while each individual corruptible thing at some time doesn't exist, there nevertheless is *no particular time* at which *no corruptible things exist*. Simply imagine a series of corruptible things, each one coming into existence, persisting for three seconds, and going out of existence. Imagine further that each corruptible thing in this series comes into existence during the 2nd second of the previous corruptible thing's lifespan. Finally, imagine this series of comings-to-be and passings-away stretches back beginninglessly (i.e. infinitely) into the past. Voila! This is a situation in which each corruptible thing is such that it doesn't exist at some time (so (1) is satisfied) but in which it is false that there is some particular time such that no corruptible thing exists at that time (so, (2) is not satisfied).[37]

A third example stems from Malpass & Morriston (2020). Suppose a hotel with an infinite amount of rooms is impossible, but that a hotel with any finite number of rooms is possible. Suppose further that there is a God who can do anything that is possible to do. We can then distinguish between the following two formulas:

(1) For each natural number n, it is possible that God creates n hotel rooms (within a hotel).
(2) It is possible that, for every natural number n, God creates n hotel rooms (within a hotel).

Claim (1) says that for *each* natural number *n*, it is possible that God made a hotel with *n* many rooms in total. Claim (2) says that it is possible that God made a hotel so big that there is a hotel room for *every* natural number *n*.

Given our assumptions at the beginning, (1) is clearly true while (2) is clearly false. God *could* make a hotel all in one go such that the hotel rooms are numbered 1 to *n*, for any *n* (claim (1)), but he *could not* make a hotel all in one go where every single natural number corresponds to a unique room (claim (2)). In slogan form, *each is possible, but not all*.

Hopefully by now you have a better grasp of scope distinctions and quantifier shifts.

Strategy #19

Show that a hypothesis or proposition entails that something is a brute (i.e. unexplained or, worse, inexplicable) fact. Though there's disagreement about whether every fact (or at least every contingent fact) has an explanation[38], there *is* agreement that (all else being equal) explanations are more probably present than not. In other words, brute, arbitrary, unexplained phenomena are unlikely or improbable. So, any hypothesis H that *entails* (or makes likely) something brute or unexplained thereby faces some evidence weighing against H.

Strategy #20

Every day we encounter explanations of various phenomena. For instance, we may see a leaf fall from a tree, or witness a dog wince, or hear a loud noise coming from the porch. Each of these call for explanations of various sorts, some of which will be much more probable than others. Which of the following, in your view, is the more probable explanation for your observation of the leaf falling?

1. (1) The leaf fell because of various natural causal factors, such as the presence of a very strong wind combined with the gravitational pull that the earth exerted on the leaf.
2. (2) There are ghostly spirits lurking in the ground that desire the nutrients in the leaf and have the power to cause the leaf to fall.

Now, we all know the former is a more plausible explanation, but it is not enough as rational thinkers merely to affirm *that* this is the case. We need to

substantiate *why* this is the case.

There are (broadly) at least five criteria for finding the best explanation in both philosophical and everyday reasoning.

Fruitfulness: has this explanation been put to the (empirical or philosophical) test and yielded results? What predictions does it make about how we would expect the world to be? Are those predictions met? Are they surprising or novel predictions?

Simplicity: the explanation that makes less claims and posits less entities is generally the better one.

In considering simplicity, take into account (i) the number of entities a hypothesis/explanation posits, (ii) the theory's ontological commitments, and (iii) the *type* of entities a hypothesis posits. Also important is the distinction between *quantitative simplicity* and *categorical* or *qualitative simplicity*. The former refers to the *number* or *quantity* of entities (or terms) within a hypothesis or theory. The latter, by contrast, refers to the number of *fundamental categories or kinds* of entities (or terms) within a hypothesis or theory.

Simplicity is principally judged by ontological commitment; theories that cover the data with fewer postulated entities are more likely to be true (all else being equal).

Testability: is the explanation open to confirmation? Is it open to falsification? Has it been confirmed? Is it unfalsifiable? Are the mechanisms specified? Are the mechanisms testable, coherent, and intelligible?

Generally speaking, falsifiable hypotheses are testable, and testable explanations tend to be the better ones. Note that this criterion or indicator of explanatory success primarily applies to *empirical* hypotheses (i.e. hypotheses amenable to scientific or experiential investigation).

Conservatism: how does the hypothesis cohere with what we already know? How well it is supported by previously established facts, evidence, and explanations?

Explanatory scope: which hypothesis sufficiently explains the largest amount of data with the most plausible and fitting explanation? Explanatory scope can be broken down into explanatory *breadth* and *depth*.

Explanatory breadth: how *much* data does the hypothesis explain? Does it unify a host of disparate phenomena into one, integrative understanding or paradigm?

Explanatory depth: how *well* does the hypothesis explain the data? Does it give us a deeper understanding of what it explains, or does it merely give a superficial understanding?

Evaluating hypotheses and theories also requires taking into account intrinsic probability, accuracy, and coherence.

Allow me to unpack these a little bit further. The best hypothesis is the one with the overall best balance of intrinsic probability and accuracy.[39]

The intrinsic probability refers to the probability[40] of the hypothesis *independent* of the evidence we have for or against it. The intrinsic probability of a hypothesis is determined by its modesty (i.e. simplicity) and coherence. Modesty is essentially a measure of how *much* the hypothesis asserts. The more a hypothesis claims, the more ways there are for it to be false and, as a result, the less likely it is to be true (prior looking at the evidence, that is). Coherence is a measure of how well the parts of a hypothesis fit together. If the different parts count against each other, the hypothesis is less coherent and thus less likely to be true.

The accuracy of a hypothesis means the degree to which its predictions correspond with reality. We measure accuracy by looking at data or evidence. Evidence is some piece of information that makes a hypothesis more probable than it would have been otherwise. Let's look at an example to get a better grasp of evidence.

Imagine you have two jars of red and blue jelly beans. In the first jar, 90% of the jellybeans are blue and the rest are red. In the second jar, 90% of the jellybeans are red and the rest are blue. Now imagine you are handed a jelly bean from one of the jars, but you don't know which jar it came from. If it's a blue bean, that's *evidence* it came from the first jar, not the second. The blue bean doesn't *disprove* that it came from the second jar because the second jar also has blue beans; nevertheless, it's *more likely* that it came from the first because there are more blue jelly beans in the first than in the second. Similarly, if it's a red bean, that's evidence it came from the second jar.

Mathematicians have a formula called Bayes' Theorem that can be used to specify the relationship between intrinsic probability, accuracy, and the overall or final probability of a hypothesis. It follows from Bayes' Theorem that a hypothesis is probably true just in case it has a greater overall balance of intrinsic probability and explanatory power than do its alternatives collectively (Lowder (2018)).

Bayesian reasoning is all about looking at competing hypotheses and asking whether we would expect the data given the hypotheses in question. If the data is expected on one hypothesis H_1 but unexpected or surprising on another hypothesis H_2, the data is evidence for H_1 vis-à-vis H_2.

One of the most powerful forms of Bayes' Theorem is known as the *odds form*. Here is its equation (don't worry, it's not as daunting as it may seem[41]):

$$\frac{P(H\,|\,E)}{P(\sim H\,|\,E)} = \frac{P(E\,|\,H)}{P(E\,|\,\sim H)} \times \frac{P(H)}{P(\sim H)}$$

Where

- E is some piece of evidence or data
- H is some hypothesis, and ~H is the denial of H

- $P(H \mid E)$ means the probability of H *given* (i.e. conditional upon) E. It refers to the posterior probability of H (i.e. posterior to a consideration of E).
- $P(E \mid H)$ means the probability of E given (i.e. conditional upon) H. It refers to the likelihood that E would obtain conditional upon the truth of H.
- $P(H)$ means the intrinsic or prior probability of H (i.e. prior to considering E)
- The left side of the equation is the *ratio of the posteriors*. The right side of the equation is the *likelihood ratio* multiplied by the *ratio of the priors*.

Example

Suppose you're a nurse screening 100 patients for Diseasitis.[42] The screening test involves a tongue depressor that usually turns black for patients who have the sickness. Suppose you know the following:

- 20% of patients in the screening population actually have Diseasitis
- Among patients with Diseasitis, 90% turn the tongue depressor black (true positives).
- However, 15% of the patients without Diseasitis will also turn the tongue depressor black (false positives).

The question now is: what is the probability that a patient with a blackened tongue depressor actually has Diseasitis?

To answer this, we can use the odds form of Bayes' Theorem. Just let H be the hypothesis that the patient actually has Diseasitis and E be the evidence of a blackened tongue depressor. With this in hand, let's reason this out.

Because 20% of the population has Diseasitis while 80% of the population doesn't, the ratio of the priors (of H to ~H) is 1:4. This is

because we start with *four times fewer people with Diseasitis as people without it*.

Then each person with the disease is 90% likely to make the depressor black, and each person without the disease is 15% likely to make the depressor black. So, the likelihood ratio is 6:1, since each person with Diseasitis is *6 times as likely* (0.90/0.15 = 6) as a person without Diseasitis to make the depressor black.

So we multiply the likelihood ratio (6:1) by the ratio of the priors (1:4) and end up with:

$$\frac{6}{1} \times \frac{1}{4} = \frac{6}{4}$$

So, the ratio of the posteriors is 6:4 or 3:2. To convert this *relative* proportion into an *absolute* probability that a random person who blackened the tongue depressor actually has Diseasitis, we calculate 3/(3+2) to see that 3/5th (60%) of the those who blacken the tongue depressor have Diseasitis. In other words, given that a random patient blackened the tongue depressor, there is a 60% chance that he or she actually has Diseasitis. Zoinks!

We'll apply the odds form of Bayes' Theorem to David Hume in a later chapter.

Strategy #21

Don't forget background rates!

> To persuade you that I am a professional archer, it is obviously insufficient to show you a bull's-eye I have made. We need to know *how many times I missed*. Getting a bull's-eye in one shot tells quite a different story than getting a bull's-eye in, say, a thousand.

Horoscopes are another common example. To properly evaluate the "evidence" adduced by believers in astrology, we need to know something else as well: how many horoscopes don't come true.

To evaluate the reliability of an argument featuring a few vivid examples, then, we need to know the ratio between the number of "hits" (so to speak) and the number of tries. It's a question of representativeness again. Are the featured examples the only ones there are? Is the rate impressively high or low? (Morrow & Weston (2015))

Suppose Jones argues as follows: 'In the second half of 2010, the University of Western Ontario did not have a single car stolen on campus. The campus police must be doing an outstanding job.'

Jones, however, is failing to consider a number of background rates relevant to his assessment. To properly justify his assessment, he would need to do the following:

Overall, we need to know how many cars there are on campus: that is, we need to know (or readily be able to calculate) the theft rate on campus. Second, we would need to know what the rate of car thefts is in the surrounding area. If there are few cars on campus or a very low theft rate in the surrounding area, then it's not as big an accomplishment if there have been no car thefts on campus. We could get even more precise here by figuring out how we're going to count the number of cars on campus. Are we looking at the number of cars on campus on, say, an average weekday morning at 10 a.m.? The average number of cars parked on campus overnight? The total number of cars parked on campus at any point during the month? All of these suggestions are ways of refining the basic idea that we need to know how many cars there are on campus (Morrow & Weston (2015)).

He should also do the following. First, look at the number of cars stolen far back into the past, such as the early 1900s. If the number has always been zero, well then there may not be anything particularly extraordinary about the current campus police. Second, look at other types of crimes – for example murders, theft, and so on. The police may be good at protecting the university from car theft, but it doesn't follow that they are doing a good job in other respects for protecting the campus. Third, look at the surrounding city/town/area/province. Maybe it is an extremely stable and safe city with low crime all around and low car theft all around. In that case, there is likely nothing special about the campus police. Fourth, look at how many cars there actually are on campus. Maybe 98% of the students and faculty bike, walk, or take the bus to classes. In that case, there will be very few cars there in the first place, meaning of course the amount of stolen cars will be incredibly small.

2.5 Tools

Finally, let's examine six tools to add to your critical thinking and philosophical reasoning toolkit.

2.5.1 *Seeing not* versus *not seeing*

A critically important distinction that arises in a whole host of philosophical debates is that between *not seeing* (where the sight in question is sight of the intellect, i.e. grasping or understanding some truth, phenomenon, etc.) versus *seeing not*. To disambiguate this from perceptual sight, the distinction can also be cast in terms of *not being aware of x* versus *being aware of not-x*, or as *not knowing p* versus *knowing not-p*.

To understand why *not seeing* neither means nor entails *seeing not*, it is useful to consider concrete examples:

- Merely from the fact that you are not aware of any neutrinos passing through you, it does not follow that you are aware that no neutrinos are passing through you;

- Merely from the fact that you are not aware that your mental states are correlated with neurophysiological states, it does not follow that you are aware that your mental states are not correlated with neurophysiological states;
- Merely from the fact that you don't see how Goldbach's conjecture[43] is impossible, it does not follow that you see that Goldbach's conjecture is not impossible; and
- Merely from the fact that person x is not aware that, upon seeing the Morning Star, he thereby sees the Evening Star, it does not follow that he is aware that, upon seeing the Morning Star, he does not thereby see the Evening Star.

Although it is a subtle distinction, it has far-ranging consequences. For instance, in debates surrounding the problem of evil, some argue that there exist certain (token or types of) evil states of affairs such that God could have no justifying reasons for allowing such states of affairs to obtain.

In arguing for this, however, some thinkers presuppose that because we are *not aware* of any possible God-justifying reasons, there must actually *be* no such God-justifying reasons. But this does not follow: Merely from the fact that we do not see God-justifying reasons, it does not follow that we see the absence of God-justifying reasons.

In cases of the *absence of sight* versus *sight of an absence*, the inference from the former to the latter seems warranted only if the following condition is met, where P is some proposition and S is some state of affairs:

> Were P to be true (or S to obtain), we *would* (likely) (i) be aware that P is true (or that S obtains) and (ii) recognize P (or S) *qua* P (or *qua* S).[44]

This condition nicely accounts for the cases in which the inference from *non-sight* to *sight-not* is clearly unsuccessful. Take, for instance, bacteria. The inference from 'I am not aware of any bacteria in my room' to 'thus, there (likely) are no bacteria in my room' is clearly quite poor. This is because, firstly, *were* there to be bacteria in your room, you *would not likely*

be aware of them. But even if you *were* aware of them (say, because you have ingested some and this is giving you a stomach ache of which you are aware), you would not be aware of them *qua* bacteria but only *qua* their manifestation in a stomach ache.

2.5.2 Misplaced modal quantifiers

Modal quantifiers, such as 'possibly' and 'necessarily', quantify over a domain of objects in possible worlds (i.e. total ways things could have been). Importantly, though, modal quantifiers are often misplaced, and knowing how to spot such misplacement is a crucial philosophical skill (which we covered in strategy #18 but expand upon here).

Consider the following claim: 'X couldn't have a cause.'

This may seem reasonably straightforward, but it is ambiguous between two possible meanings:

(1) There does not exist a C such that C possibly causes X. (Else: for all existing C, it is impossible that C causes X)
(2) It is not possible that there exists a C such that C causes X (Else: it is impossible that for any C, C causes X).

The former states that, *of the actually existing things*, none of them can cause X. Importantly, though, this still allows that X possibly has a cause – so long as the cause is one that *could* exist but does not actually exist. After all, (1) only quantifies over the actually existent things, and hence does not quantify over something that *could* exist but does not actually exist. The latter, by contrast, says that it is simply impossible for anything – whether actually existent or merely possibly existent – to cause X.

Relatedly, it is crucial to disambiguate between:

(1) Necessarily, if p then q.
(2) If p, then necessarily q.

For instance, if I know that Jones will get married tomorrow, then it logically follows that Eli will get married tomorrow. In this case, knowledge of x's being the case entails x is the case.

But, clearly, it is not *necessary* that Jones gets married tomorrow. After all, Jones could come down with a high fever, postponing the wedding until next week. We therefore have, in this case:

(1) Necessarily, if I know that Jones will get married tomorrow, then Jones will get married tomorrow.

But this neither means nor entails that, if I know Jones will get married tomorrow, then Jones *must necessarily* get married tomorrow. In other words, the following is false:

(2) If I know that Jones will get married tomorrow, then necessarily Jones will get married tomorrow.

This becomes especially important in discussions concerning logical fatalism, free will, omniscience, open future, and so on.

In addition to the aforementioned distinction, a number of other important distinctions and points come into play. Here is a short list of them for you to add to your toolkit:

- Just because each individual thing in a domain could fail to be, it doesn't follow that there could be a state in which there fails to be all things in the domain (together). This is a subtle form of the quantifier shift fallacy: Just because each individual student has a counselor, it doesn't follow that there is a counselor for all the students (together).

This is relevant in a number of ways. Some thinkers mistakenly reason that, because there could not possibly be nothing, there must exist a necessary being. But notice that this is subtly fallacious. The impossibility of there being nothing is equivalent to the necessity of there being something. But the necessity of there being *something or other* does not entail that there

exists *some specific, particular thing* that necessarily exists. Similarly, even if every individual object in reality is contingent, it does not follow that, possibly, there are zero objects in reality. This is because it could still be the case that, necessarily, there is *some contingent thing or other* despite the fact that, for any given *particular contingent thing*, that thing could fail to exist.

- Just because each individual thing in a domain is necessarily and essentially non-F, it doesn't follow that all the things – together – in the domain are necessarily and essentially non-F.

This is important with respect to to materialism and dualism. Some dualists will argue that each neural subsystem is essentially and necessarily non-conscious, in which case adding more and more subsystems together could only beget further states which lack consciousness. But this inference is not necessarily truth-preserving. For instance, when adding two chemicals together that are essentially and necessarily colorless, their interactions can nevertheless give rise to something with a vibrant red color.

- Finally, if two things are such that it is impossible for one to exist without the other, it does not follow that they are identical. For instance, the property *having a radius* cannot exist without the property *having a circumference* (and vice versa); nevertheless, they are clearly not identical properties.

2.5.3 Intensional and extensional contexts

A linguistic context is intensional precisely when it is not extensional. An extensional context is (for our purposes) present when the following condition is met:

(1) Necessarily, inter-substituted co-referring expressions preserve truth.

Condition (1) applies to scenarios in which an object has two names, N_1 and N_2. In such scenarios, one can formulate a true proposition about the object using N_1. When condition (1) is met, if you substitute N_2 for N_1 into the proposition, the proposition cannot become false. In other words, such a substitution is truth-preserving. Because the object to which each name refers is identical, the words used to denote it simply don't matter to proposition's truth (in extensional contexts).

But (1) is not always met, and when it isn't, the linguistic context is *intensional*. In intensional contexts, co-referring expressions cannot be substituted for one another without potentially changing the proposition's truth value. The failure of (1) in intensional contexts derives from the fact that a single object can be presented to a thinking subject under different *aspects* or *clusters of descriptions*. Under such conditions, the subject in question may be unaware that the extramental *referent* of such clusters of descriptions is the same in each case.

In summary, an intentional object (else: linguistic expression) construed *extensionally* is the intentional object's extramental *referent*[45], whereas an intentional object construed *intensionally* is the mind-dependent cluster of descriptions that the thinking subject associates with an object. This will become especially important in the chapter on the nature of mind.

2.5.4 The A-series and B-series

The terminology of the A-series and B-series (and the corresponding A- and B-theories of time) for characterizing time is due to John McTaggart. This is an invaluable tool in your philosophical toolkit since many arguments may implicitly presuppose the truth or falsity of one such theory of time.

The A-series is a group of succeeding temporal events that are past, present, and/or future. This series is *dynamic* because the tensed operators of past, present, and future are always changing depending upon what time is present. These tensed operators aren't mere relations – they're absolute.

The B-series is an extended group of temporal events that stand in the relations of *earlier than*, *later than*, or *simultaneous with*. This series is *static* because the tenseless operators previously articulated are unchanging: when x is earlier than y, x is always earlier than y and cannot transition to becoming later than y.

Roughly, A-theories of time hold that:

- The A-series is the fundamental characterization of time.
- Tensed operators pick out objective features of reality.
- Reality is dynamic and changing – temporal becoming is a real and objective feature of reality.
- Not all times are equally actual.
- Past, present, and future are absolute and objective determinations, not relative to reference frame, observer, perspective, other times, and so on.

B-theories stand in opposition to A-theories in that B-theories hold (roughly) that:

- The B-series is the fundamental characterization of time.
- Tensed operators do not pick out objective features of reality.
- Reality is static and temporally unchanging – temporal becoming and temporal passage are neither real nor objective features of reality.
- All times are equally actual.
- Past, present, and future are neither absolute nor objective determinations.

Some important arguments thinkers have leveled in favor of A-theory:

- The argument from our phenomenological experience of temporal becoming

- - Our experiences (both perceptual and phenomenological) are pervasively infused with objective, dynamic flow and temporal passage.
  - The argument from the inadequacy of B-theoretic accounts of change
    - Change involves the actualization of things which are not-yet existent, i.e. potentially existent – in which case, if all of temporal reality were equally actual, there would be no change within temporal reality (which is absurd).
  - The argument from the ineliminability of tense
    - If B-theory is true, then there must be adequate translations of tensed language into tenseless language; but no such adequate translations could exist.

Some important arguments thinkers have leveled in favor of B-theory:

- Special and general relativity indicate that (i) there cannot be absolute simultaneity, (ii) time is relative to frame of reference, and (iii) space is inextricably bound up with time – time and space form a single four-dimensional spacetime block.
  - As one frame of reference approaches the speed of light relative to another frame of reference, time in the former frame of reference passes much more slowly.
- The argument from successful reference to past and future times.
  - We successfully refer to things in the past and future; but one cannot successfully refer to x unless x actually exists.
- The argument from time-indexed truthmakers
  - There are truths about the past and future; but truths require actually existent states of affairs serving as truth-makers; and the only possible truth-makers for time-indexed propositions are the time-indexed states of affairs themselves.

- The argument from the indeterminacy of experience between the truth of A and B theories
  - The order and intrinsic character of events in the A-series and B-series are identical, in which case the causal order between the two series is identical, in which case (since experience of x requires a causal connection to x) experience would be identical in either the A-series or B-series.
- The argument from causation
  - Past temporal states stand in the relation of causation to present things; but only actually existent things can serve as causes; hence, past temporal states must actually exist.
- The argument from existence-entailing relations
  - The past and future stand in all sorts of relations to the present; but x and y can only stand in a relation if x and y actually exist; hence, the past and future actually exist.

Once again, my purpose is only to give you a rough picture of the conceptual landscape. I am not endorsing or denying any of these specific arguments for either view of time.

2.5.5 Different types of possibility and necessity

It is crucial to disambiguate what we mean when we say that x is possible or not possible. Here are various ways to disambiguate it:

(1) Conceptually possible

Something's being conceptually possible means that it, as a concept in itself, has no internal contradiction of incoherencies. So, arguably Plato's Realm of Forms is conceptually possible, since there seems to be nothing utterly incoherent about the concept of a non-spatiotemporal, acausal reality in which abstracta exist.

(2) Logically possible

Something's being logically possible means that its negation cannot be derived from (i.e. is not entailed by) the true logical axioms. For instance, arguably logical axioms alone do not entail that 'the number 32 cannot produce a badminton player', in which case 32's producing a badminton player may be logically possible even though it is not metaphysically possible. This example illustrates that, when we talk about whether a claim is logically possible, we consider only the *logical* features of the sentence and not the nature of the objects or properties to which the claim refers.

(3) Metaphysically possible

Metaphysical possibility is a subset of logical possibility and is more stringent than logical possibility. Something's being metaphysically possible means that it exists in at least one possible world, where a possible world is just a complete way the actual world could be (else: could have been).

That in virtue of which worlds are (metaphysically) *possible* depends on the correct modal metaphysics. According to Aristotelian accounts of modality, x is *possible* provided that x is either actual or potential, where x is *potential* provided that there is some actual thing with the causal power to initiate a causal chain leading to x's actuality. According to Platonic accounts, x is possible in virtue of being a member of a maximal consistent (else: compossible) set or collection of propositions.

Leibnizian accounts hold that possible worlds are maximal self-consistent ideas or concepts in the divine mind. According to Lewisian (extreme modal realist) accounts, x is possible in virtue of being true in an actually existent, spatiotemporally related concrete world, where actuality is indexical; all possible worlds genuinely exist and are on ontological par with one another in having the same modal status.

Importantly, metaphysical possibilities have to be consistent not only with the logical landscape, but also with (i) the necessary truths of reason and reality (e.g. if container C_1 contains a container, C_2, then C_1 contains

whatever C_2 contains), and (ii) the *nature* of substances, events, properties, and so on.

Examples of metaphysical impossibilities include:

Something cannot be red and green all over. Solely in terms of logical features this sentence is perfectly consistent: 'It's impossible for something to have property R and property G on all its color-reflective parts.' Hence, if the proposition is true, it has something to due with the nature of color properties. Likewise, *prime numbers cannot produce prime ministers*.

(4) Epistemically possible

Something's being epistemically possible means that, for all we know (or justifiably accept), x could be true; x is not known to be false or impossible.

For instance, it is epistemically possible that Goldbach's conjecture is true, but it is also epistemically possible that Goldbach's conjecture is false. Nevertheless, apart from our justified acceptances, it is either necessarily true or necessarily false. Similarly, I currently don't know whether it is raining in London. So, epistemically, it is possible (given my justified acceptances) that it rains in London, and it is possible that it does not.

(5) Temporally possible

Something's being temporally possible is essentially a matter of time-indexed metaphysical possibility. In other words, something is temporally possible provided it is metaphysically possible when indexed or specified at a particular time (or collection of times). It is essentially *tensed* metaphysical possibility (i.e. it includes reference to one or more times)

Some examples of temporal modality:

- X necessarily holds at the next temporal state of reality.

- X eventually cannot fail to hold (somewhere on the subsequent temporal path of reality).
- X cannot fail to hold on the entirety of a potential subsequent temporal path of reality.
- X cannot fail to hold on the entirety of every single potential subsequent temporal path of reality.
- There exists at least one temporal path reality *could* take starting from the present moment where X holds.
- Possibly (or necessarily), it will sometimes be the case that X.
- Possibly (or necessarily), it will always be the case that X.
- Possibly (or necessarily), it was sometime the case that X.
- Possibly (or necessarily), it has always been the case that X

(6) Physically possible

Something's being physically possible means that it is consistent with the actual laws of nature. Traveling faster than the speed of light, for instance, is physically impossible, whereas having 20 children is physically possible.

2.5.6 Putting the grounded before the ground

Putting the grounded before the ground is a version of putting the cart before the horse. It consists in mistakenly thinking that X is grounded in Y, whereas Y is, in reality, grounded in X.

For instance, even if it's already true that, say, tomorrow I will eat chicken for lunch, it's not *in virtue* of this proposition's truth that I eat the chicken. Rather, the proposition is true *in virtue of my eating chicken*. So, although I may not 'change' the truth value of the proposition (which may give the illusion of my not grounding it), we still have counterfactual dependence: *had* I chosen instead to eat fish, the proposition *would have* been false, and it is presumably *my eating chicken* (or at least the *fact* that I will eat chicken) that makes it be the case that the proposition is true.

Another example concerns the argument that B-theories of time entail the absence of free will. Arguably, this puts the grounded before the ground.

My decision to *phi* in the future in the B-series is not grounded or true in virtue of its being 'already actual' in the future; rather, it's the other way around. My decision to *phi* in the future is actual precisely in virtue of *my choice to phi*. The event is actual in virtue of my *choosing* it to be actual; I don't choose it to be actual in virtue of its *actuality*. Don't put the cart before the horse.

So, even if the future is already actual, it doesn't follow that I cannot do otherwise, since the actual status of the future event is grounded in my choice rather than vice versa. The future would have different actual features had I chosen non-phi, meaning that the future could have been different even if it is already actual.

2.5.7 Formal logic

The various rules of inference are essential tools for your critical thinking toolkit. While the following list is not exhaustive, it should be reasonably representative.[46] I strongly recommend working through and understanding each one in turn.

Key:

- p, q, r, and s refer to propositions (i.e. declarative statements that are either true or false)
- → refers to material implication. In other words, (p → q) means 'if p is true, then q is true' or 'if p, then q' or 'p implies q' or 'p entails q'
- ∴ refers to 'therefore'
- ∨ refers to 'or' (this is the inclusive or – (p ∨ q) means p or q or both)
- ∧ refers to 'and'
- ~ refers to negation (so, ~p means 'not p' or 'it is not the case that p')

Modus Ponens

(p → q)

p

∴ q

If p then q; p; therefore q

Modus Tollens

(p → q)

~q

∴ ~p

If p then q; not q; therefore not p

Hypothetical Syllogism

(p → q)

(q → r)

∴ (p → r)

If p then q; if q then r; therefore, if p then r

Disjunctive Syllogism

(p ∨ q)

~p

∴ q

Either p or q; not p; therefore, q

Constructive Dilemma

(p → q)

(r → s)

(p ∨ r)

∴ (q ∨ s)

If p then q; and if r then s; but either p or r; therefore either q or s

Destructive Dilemma

$(p \rightarrow q)$
$(r \rightarrow s)$
$(\sim q \lor \sim s)$
$\therefore (\sim p \lor \sim r)$

If p then q; and if r then s; but either not q or not s; therefore either not p or not r

Simplification

$(p \land q)$
$\therefore p$

p and q are true; therefore p is true

Conjunction

p, q
$\therefore (p \land q)$

p and q are true separately; therefore they are true conjointly

Addition

p
$\therefore (p \lor q)$

p is true; therefore the disjunction (p or q) is true

Composition

$(p \rightarrow q)$
$(p \rightarrow r)$
$\therefore (p \rightarrow (q \land r))$

if p then q; and if p then r; therefore if p is true then q and r are true

De Morgan's Theorem (1)

~(p ∧ q)
∴ (~p ∨ ~q)

The negation of (p and q) is equivalent to (not p or not q)

De Morgan's Theorem (2)

~(p ∨ q)
∴ (~p ∧ ~q)

The negation of (p or q) is equivalent to (not p and not q)

Commutation (1)

(p ∨ q)
∴ (q ∨ p)

(p or q) is equivalent to (q or p)

Commutation (2)

(p ∧ q)
∴ (q ∧ p)

(p and q) is equivalent to (q and p)

Association (1)

(p ∨ (q ∨ r))
∴ ((p ∨ q) ∨ r)

p or (q or r) is equivalent to (p or q) or r

Association (2)

(p ∧ (q ∧ r))

∴ ((p ∧ q) ∧ r)

p and (q and r) is equivalent to (p and q) and r

Distribution (1)

(p ∧ (q ∨ r))
∴ ((p ∧ q) ∨ (p ∧ r))

p and (q or r) is equivalent to (p and q) or (p and r)

Distribution (2)

(p ∨ (q ∧ r))
∴ ((p ∨ q) ∧ (p ∨ r))

p or (q and r) is equivalent to (p or q) and (p or r)

Double Negation

p
∴ ~~p

p is equivalent to the negation of not p

Transposition

(p → q)
∴ (~q → ~p)

if p then q is equivalent to if not q then not p

Material Implication

(p → q)
∴ (~p ∨ q)

if p then q is equivalent to not p or q

Exportation

$((p \land q) \rightarrow r)$
$\therefore (p \rightarrow (q \rightarrow r))$

From (if p and q are true then r is true) we can infer (if q is true then r is true, if p is true)

Importation

$(p \rightarrow (q \rightarrow r))$
$\therefore ((p \land q) \rightarrow r)$

From (if p then (if q then r)), we can infer (if (p and q), then r)

Tautology (1)

p
$\therefore (p \lor p)$

p is true is equivalent to p is true or p is true

Tautology (2)

p
$\therefore (p \land p)$

p is true is equivalent to p is true and p is true

Law of Excluded Middle

$\therefore (p \lor \sim p)$

Either p is true or p is not true

2.6 And so it begins…

Whew! That ends our tour of the some of the methods and tools required for critical thinking in philosophy (and everyday life, for that matter). The aim of this chapter is to *serve* you – to equip you with powerful tools of rational thinking and to give you a greater sense of the beauty and majesty of reason.

But without practicing and implementing these various methods and tools, *they won't stick*. That's why the next three chapters will be 'case studies' (as it were) wherein we develop these skills further. Chapter three will focus in particular on scientism – the view that science is the only path to knowledge. Chapter four will focus on a Bayesian argument against Humeanism with respect to laws of nature. Finally, chapter five will evaluate a number of different philosophical accounts of the nature of the mind.

Once again, treasures of truth await…

Chapter 3: Scientism

1 Introduction

Science is beautiful. Ranging from the subatomic realm to the astronomical heavens, scientific findings are truly stunning. Science has illuminated the neurophysiological processes underlying your perception of the words on this page no less than it has shed brilliant light on the formation and evolution of galaxies. Biology has allowed us not only to understand the complex world of cells, organisms, and populations, but it has also revealed the evolutionary processes by which such things came to be in the first place. Chemistry, like biology, has filled libraries with knowledge of the natural world ranging from the buzzing world of electrons to the synthesis of medicinal compounds that have saved hundreds of millions of lives. Physics, taken by many to be the apex of natural science, has elevated human knowledge to unforeseen heights. Magnificently capturing the mathematical structures governing all physical bodies, physics has unified the smallest particles and largest galaxies under elegant mathematical laws whose intelligibility to the intellect is awe-inspiring.

In light of this immense scientific success, we are naturally prompted to wonder: does science capture all there is to reality? Is scientific knowledge the only genuine knowledge that exists? Why is science so successful in both its predictions and its technological development? It is in answering these questions that some thinkers are led to a position known as *scientism*. Apart from being a fascinating topic in its own right, scientism will also provide us with an excellent opportunity to apply the many things we have learned in our journey thus far.

Scientism is, in essence, the view that natural science alone provides genuine knowledge of reality. Empirical methodology and scientific experiment, in other words, are the sole means by which knowledge is produced. Philosopher and proponent of scientism Alex Rosenberg defines scientism as "the conviction that the methods of science are the only

reliable ways to secure knowledge of anything; that science's description of the world is correct in its fundamentals" (Rosenberg (2011)). He continues, stating that "being scientistic just means treating science as our exclusive guide to reality and to nature." It should be noted that by *science* what we mean is the description, investigation, and explanation of the physical world that is principally based on the deliverances, whether direct or indirect, of sensory experience. Science, so understood, crucially involves observation and testable experimentation and is paradigmatically embodied in physics, chemistry, and biology. Science is in the business, then, of yielding conclusions as to the nature, behavior, structure, and causal relationships of natural phenomena.

As understood in the aforementioned sense, scientism maintains that science is the *only* reliable means of generating knowledge, i.e. of ascertaining truth. Though related to scientism so defined, the question of whether science is the *best* means of generating knowledge is not what is principally at issue here, since the mere proposal that science is the *best* means of producing knowledge is entirely consistent with *other* disciplines being reliable means of producing knowledge as well.

With the definitions of both *science* and *scientism* out of the way, we can turn now to an analysis and evaluation of scientism. Earlier we saw that the success of science poses a number of questions concerning its scope in describing, investigating, and explaining reality. More specifically, we briefly considered the question as to whether science captures all there is to reality. Because scientism answers this question in the affirmative, it is worth beginning our analysis with the motivations behind scientism. What prompts many thinkers to adopt scientism?

2 Motivations behind scientism

2.1 Success of science

Science has been a story of unrelenting success both in terms of its confirmed predictions and its technological progress. The list of technological developments resulting from science seems endless:

smartphones, computers, solar panels, prosthetic limbs, rocket ships, cars, and so on.

Science's success, though, extends far beyond mere technological progress. Its explanatory success in relation to natural phenomena is truly unparalleled, broadening and deepening our understanding of particle interactions, chemical reactions, biological and metabolic processes, and cosmic evolution. Its predictive success, moreover, is no less remarkable than its explanatory and technological success. Among its triumphs are predicting the development, spread, and treatment of diseases, the positions and interactions of heavenly bodies, the existence of intermediate fossil forms in specific sedimentary rock strata, weather and climatological patterns, and so on. Overall, then, the expansive accomplishments of science seem undeniable.

Alex Rosenberg, reflecting upon this arresting success, mounts an argument in favor of scientism. He writes:

> And it's not just the correctness of the predictions and the reliability of technology that requires us to place our confidence in physics' description of reality. Because physics' predictions are so accurate, the methods that produced the description must be equally reliable. Otherwise, our technological powers would be a miracle. We have the best of reasons to believe that the methods of physics—combining controlled experiment and careful observation with mainly mathematical requirements on the shape theories can take—are the right ones for acquiring all knowledge. Carving out some area of 'inquiry' or 'belief' as exempt from exploration by the methods of physics is special pleading or self-deception…

> The phenomenal accuracy of its prediction, the unimaginable power of its technological application, and the breathtaking extent and detail of its explanations are powerful reasons to believe that physics is the whole truth about reality (Rosenberg (2011)).

The idea, then, is that the unparalleled success of science (understood in terms of predictive accuracy, technological application, and explanatory breadth and depth), and in particular physics, compared to any other discipline gives us a strong reason to believe that science alone is the sole means for attaining truth. Although Rosenberg is more restrictive in his scientism than other scientistically inclined thinkers (holding that physics alone is the sole epistemological guide to reality as opposed to physics in conjunction with chemistry, biology, etc.), the spirit of his argument captures the general thrust of the arguments cited by proponents of scientism.

2.2 Verificationism

Science's success, although perhaps the most powerful argument for scientism, is by no means its sole motivation. A second motivation stems from *verificationism*, a theory of meaning associated with the logical positivism that flourished in the early twentieth century. The logical positivists, impressed by the exactness and success of science, logic, and mathematics, sought to reform philosophy principally by means of verificationism. AJ Ayer, one of the foremost philosophers in the movement, stated his aim as follows: "to establish beyond question what should be the purpose and method of a philosophical inquiry" (Ayer (1936)).

In essence, verificationism is a theory of meaning that seeks to demarcate statements that are meaningful from statements that are meaningless. Its core thesis can be captured by what is known as the *verification principle* (VP), according to which a statement is meaningful if and only if it is either analytic or empirically verifiable. An analytic statement is a statement that's true or false solely in virtue of the meaning of its terms, such as the statement that *all bachelors are unmarried*. By contrast, the statement that *all bachelors are above six feet tall* is synthetic, since *being above six feet tall* is not contained within the concept *bachelor*. A synthetic statement, therefore, is true or false in virtue of the way the world actually is rather than solely in virtue of the meanings of its terms.

A statement is empirically verifiable provided there is some set of observations that would establish its truth. Consider again the statement that *all bachelors are above six feet tall*. At least in principle, there is some set of observations that would establish its truth. For instance, if we were to gather all the bachelors together and make measurements and observations of each of their heights, and if each of their heights turned out greater than six feet tall, we would know the statement is true. For any given empirically verifiable statement, the set of observations that would establish its truth is known as its *verification condition*. Verificationism (at least in one of its forms) holds, moreover, that the meaning of a synthetic statement *just is* its verification condition. From this, it follows that any synthetic statement which lacks a verification condition (i.e. a synthetic statement for which there exist no observations that could establish its truth) is meaningless.

Verificationism's relation to scientism should, at this point, be evident. Verificationism, I shall show, entails scientism. The following argument for this conclusion will solely consider synthetic statements (referred to as 'statements' in the argument), since analytic statements are trivial in the sense that they are true solely in virtue of definitions. Scientism, though, principally aims to establish that the only way to gain non-trivial knowledge about the *mind-independent world*, as opposed to knowledge of our definitions and concepts, is by means of science. The restriction of scope only to synthetic statements is thus justified.

Consider that if verificationism is true, then all statements that lack empirical observations that could establish their truth are meaningless. Now, scientific statements are, by their nature, statements that are amenable to empirical observation. In other words, scientific statements have empirical observations that could establish their truth. For the proponent of scientism, moreover, any statement which is amenable to empirical observation and experiment (i.e. any statement with a set of observations that could establish its truth) is a scientific statement. This entails, by contraposition, that a *non-scientific* statement is a statement that is *not* amenable to empirical observation (i.e. it *lacks* a set of observations that could establish its truth). Since, according to verificationism, any statement that *lacks* a set of observations that could establish its truth is meaningless,

it follows that *non-scientific* statements are meaningless. And since any statement that can be knowledge requires that the statement be meaningful, it follows that *non-scientific* statements cannot be knowledge. In other words, a statement can be knowledge only if it is a *scientific* statement, which is precisely the central claim of scientism. We can see, therefore, that verificationism, if true, establishes the truth of scientism. Why, though, should we believe in verificationism?

The central argument supporting verificationism is that statements make claims about the external world, and the only access we have to the external world is by means of our sensory experience. Hence, if there are not even potential experiences we could have that would show a statement, S, about the external world to be true (i.e. if there were no observations that could establish its truth), there is not even a potential that we have access to the state of the world affirmed by S, in which case S is totally beyond our cognitive capacities and our grasp of meanings – it lacks meaning (so the argument goes).

Consider, for instance, the statement that *there is a unicorn in my garden*. According to the verificationist, we need to ask: how could the truth of this statement affect our experience? Well, if we went to my garden, we would see the unicorn! The truth of this statement would *make a difference*. Consider now, though, the statement that *there is an invisible, intangible, colorless, odorless, inaudible, and completely undetectable unicorn in my garden*. Superficially, this statement *seems* meaningful. But, for the verificationist, it is meaningless since it *makes no difference* – our experience of the world would be exactly the same regardless of its truth value. What would it even mean for a unicorn to *exist* but yet not be detectable in any way, shape, or form? The verificationist capitalizes on the intuition (however faint or strong) behind this question.

2.3 The nature of science: objectivity and agreement

A third argument in favor of scientism is that science is characterized primarily by objectivity and rigor while philosophy and other disciplines are fraught with subjectivity. Science, in other words, reduces if not

eliminates sociological, personal, psychological, economic, cultural, and political factors from its methodology and findings, while philosophy and other disciplines cannot escape such personal biases and prejudices. Any philosophical position, for instance, is mired in the subjective and personal preferences, beliefs, intuitions, and history of the individual philosopher who proposed it. Science, by contrast, eliminates such biases by means of (i) its systematic peer review process (wherein scientists of all different backgrounds evaluate the assumptions, aims, methodology, data, results, conclusions, and explanations of scientific research) and (ii) its confinement to observable, quantifiable, experimental procedures that are accessible to and repeatable by anyone (at least in principle). For this reason, only science (it is argued) can give genuine, objective knowledge about reality.

Moreover, science – since it helps itself to a wide range of experimental instruments and analytical techniques – tends to be more technically assisted and meticulously procured than other disciplines that draw (at least in part) from more mundane ordinary experience. It seems, therefore, that empirical observation, and by extension science itself, is much more trustworthy than other disciplines that rely on mundane experience combined with other non-observational and non-experimental ways of knowing.

Science is objective in yet another sense, whereas (it is claimed) other disciplines such as philosophy are not. For consider that science is grounded in observations that different people can make of the exact same phenomena. The objects of scientific investigation, in other words, are *publically accessible* to a range of people by means of observation and sense experience. A whole host of different scientists and laymen alike can read the same measuring balance, thermometer, or pressure gauge, and this in turn removes the subjectivity of the human observers from their observations. To Smith, substance X may feel warmer than substance Y, whereas to Jones substance Y feels warmer than substance X. Science allows an objective measurement of the actual temperatures of substances X and Y, thereby foregoing the subjectivity of Smith and Jones. In contrast to the public accessibility and third-person nature of empirical phenomena and scientific investigation, philosophy and other disciplines (it is argued) lack

the possibility of independent verification and examination. The objectivity of science in this regard gives yet another reason to affirm it is the sole reliable guide to truth.

The nature of scientific findings as exact and mathematically precise also facilitates *agreement* among scientists, whereas the inexact and imprecise nature of philosophy and other disciplines facilitates massive disagreement. Although disagreements occasionally crop up in science, there is nevertheless substantial general agreement on its methods, principles, and findings. Surely if philosophy and other disciplines produced genuine knowledge about reality, those who practice such disciplines would be able to convince each other and form a consensus. Yet this is by no means the case; in fact, there is widespread disagreement not only about their conclusions, but also about their methodology and starting points.

With the explication and motivation behind scientism out of the way, we shall turn now to its analysis and evaluation. In doing so, we will implement the methods and skills discussed thus far in the book and, in the process, further develop your critical thinking skills.

3 Responses

3.1 The success of science: a response

There is, of course, no denying that the predictive and technological success of science is unrivaled. But it simply doesn't follow from this fact that science is the exclusive arbiter of what is true and what is false. From the mere fact that science has had tremendous success, it simply doesn't follow that science is the sole reliable guide to truth. The argument really amounts to the following:

1. The predictive success and technological applications of science are unrivaled by all other disciplines.
2. Therefore, what science reveals to us is all that exists (else: all that can be known).

As Feser (2017) points out, to argue as such is tantamount to arguing as follows:

1. Metal detectors have had far greater success in finding coins and other metallic objects in more places than any other method has.
2. Therefore, what metal detectors reveal to us (coins and other metallic objects) is probably all that exists (else: all that can be known).

Since this latter argument is absurd, and since it parallels the former argument, it follows that the former argument is absurd as well.[47] The key thing to realize is that metal detectors are *fine-tuned* (as it were) to those parts of reality amenable to metal detection in the first place. In other words, metal detectors are, by their nature, exclusively attuned to capturing things which permit the passage of electric current by exposure to magnetic fields. As Feser points out, though, "however well they perform this task – indeed, even if they succeeded on every single occasion they were deployed – that simply wouldn't make it even probable that there are no aspects of the natural world other than the ones they are sensitive to" (Feser (2017)). Likewise, scientific investigation is *fine-tuned* to capturing the portions of reality that are amenable to empirical detection, mathematical modeling, and so on. While this does in fact contribute to science's tremendous success, it is simply a non-sequitur to conclude that there do not exist *any other aspects of reality*.

Now, proponents of scientism are likely to object that the parallel argument is *relevantly dissimilar* (to use a phrase introduced in an earlier chapter) to the argument from the success of science since metal detectors can only detect a *portion* of reality while science, by contrast, detects the *entirety* of it. But this reply begs the question at issue, for whether science really describes the entirety of reality is precisely the question at issue. This cannot amount, then, to a symmetry breaker between the cases of scientism and metal detectors.

It also won't do to challenge philosophy and other disciplines to display *their* technological and predictive success, and conclude from the absence of a list to the unreliability of such disciplines. This would be no less absurd than demanding a "list of the metal-detecting successes of gardening, cooking, and painting, and then concluding from the fact that no such list is forthcoming that spades, spatulas, and paintbrushes are all useless and ought to be discarded and replaced with metal detectors" (Feser (2017)).

The underlying point here is that the *standards of success* are different for different disciplines. Success for a historian will consist in, say, illuminating the causes, nature, effects, and significance of a historical event or person. To demand that the standards of success for science must apply to other disciplines for their reliability and truth-directedness, then, is to *implicitly presuppose* that success *as such* can only consist in *scientific success*. But again, that is the very issue in question, and hence such a demand is simply question-begging.

3.2 Contra verificationism

Although logical positivism (and by extension verificationism) maintained its ascendency throughout the early and mid twentieth century, it foundered on a number of critical flaws that led to its ultimate demise. For starters, consider again the verification principle: a statement is meaningful if and only if it is either analytic or empirically verifiable. This principle itself, though, is neither analytic nor empirically verifiable, and hence according to its own criterion it is meaningless. One cannot *observe*, via one's senses, the truth of the verification principle. There is thus no set of observations which could establish its truth. But it is not analytic either, since it certainly is not definitionally and trivially true. The principle is thus self-defeating insofar as it entails its own meaninglessness.

While some have tried to defend the VP from self-refutation on the grounds that it ought to be considered a convention to be adopted on the basis of its utility, this won't do. First, "no non-verificationist has any reason to adopt this convention, and the claims that the principle has greater 'utility' than any alternative… simply begs the question" (Feser (2019)).

Statements and questions not captured within the VP's stringent criterion are taken by non-verificationists to be perfectly legitimate, meaningful, and answerable at least in principle. Such statements, in turn, possess utility in virtue of such features (since, presumably, the only basis for a statement or question's utterly lacking in utility would be its illegitimacy, meaninglessness, and/or unanswerability). Thus, to claim that the VP ought to be accepted on the basis of utility is merely to *presuppose* the non-utility and thus illegitimacy, meaninglessness, or unanswerability of statements not captured within the VP – but again, *this is the very issue in question*. This defense, then, is simply question-begging.

A second problem with this response is that to conceive of the VP and alternative conditions for meaning as mere conventions implies that *meaningfulness itself* is a mere convention. And since truth or falsity presuppose meaningfulness, rendering meaningfulness a mere convention entails that *truth and falsity* are mere conventions. But this is clearly absurd, as it entails a self-undermining relativism about truth and falsity.

3.2.1 Counter-examples

We can also use the technique of *finding counterexamples* in order to evaluate this particular argument for scientism. If, for example, we could find instances where a statement is clearly meaningful but yet is neither analytic nor empirically verifiable, we will have falsified the VP and, as a result, blocked the inference from the truth of VP to the truth of scientism. And, indeed, such statements do exist.

Take, for instance, statements about the past and future. Strictly speaking, we cannot observe future or past states of affairs, and thus we cannot observationally or empirically verify them. VP thus entails that such statements are meaningless, which seems quite obviously false. One might respond that we need only require that there is at least *in principle* some set of observations that someone could make *in the future* to verify statements about the future (a corollary principle may be adduced for past events as well). This, too, will not do. To borrow an example from Feser (2019), consider the statement that "there will be no observers in existence after the

year 3000". Clearly, to observationally verify this statement presupposes the existence of one or more observers, in which case such a statement cannot even in principle be observationally verified. But yet it is quite clearly meaningful!

3.2.2 Verificationism undermines science: mental states

Most fatally, though, is that verificationism *undermines* science itself in at least three ways. For one thing, statements concerning the private, subjective, first-person mental states of others are, strictly speaking, rendered meaningless. This is because such mental states, by their nature, are not susceptible to observational verification (as I cannot, for instance, observe your private mental states to which you alone have privileged access). This result alone is obviously false on its face, but its import cuts much deeper. It would undermine science itself insofar as scientific knowledge is dependent upon the sensory experiences and thinking processes (mental states, that is) of other scientists (Feser (2019)).

3.2.3 Verificationism undermines science: laws of nature

A second manner in which verificationism undermines science is that it entails that statements reporting laws of nature are, strictly speaking, meaningless. Laws of nature are universal statements of how natural objects can and will behave. Crucially, though, such statements could never be verified. Any number of finite observations could not establish a truth with universal import that covers an infinite amount of potential natural objects and events.

The verificationist may respond by allowing the possibility of verification through observations *making a statement's truth probable*. Although such an increase in probability could not *conclusively* observationally verify a statement, we need not be so restrictive (it is argued) about our verification conditions. The cost of such a response, though, is that it opens the door to things verificationism sought to avoid in the first place, since metaphysical, ethical, and theological conclusions are

often probabilistically but not conclusively supported by empirical observations.

Remember, further, that the meaning of a synthetic statement for the verificationist *just is* its verification condition(s). Thus, if the verification conditions are based on a finite number of limited observations that only make probable the truth of a universal law, it seems that the law itself would be stripped of its universality and would only be able to apply to those *observed instances* of the regularity. To go *beyond* such instances and infer the universality of the law requires going *beyond* the very verification conditions that lend probabilistic support to the law. Since that which observationally verifies (in the loose, probabilistic sense) the universal law is itself *not universal* and only consists in as-yet-experienced observed regularities (finite in number), it follows that the *verification condition* for a universal law is itself not universal. And since, per verificationism, the meaning of a synthetic statement *just is* its verification condition, it follows that the meaning of a universal law is itself not universal – which is absurd. By definition, a universal law has a meaning which covers a potential infinity of cases, *universally*.

Probabilistic inferences

Moreover, even allowing that mere *probabilistic* observational verification justifies inferring a universal statement, we need to countenance the truth and hence meaningfulness of statements which cannot be observationally verified in order to make probabilistic inferences. For instance, to infer *from* a host of particular observations that a given natural phenomenon or set of phenomena act in accordance with a given law of nature *to* the universal truth of the law of nature requires a premise roughly like *if it is the case that a significant number of instances regularly follow a purported law of nature N, then, probably, N is a universally true law*. Call this premise 'P'.

The problem is that P renders verificationism circular in a subtle but insidious way. For, consider that P is motivated by the thought that an observed phenomenon's regularly acting in accordance with a purported

universal law *raises the probability of* the purported universal law being universally true. But to claim that an observed regularity raises the chance of its being a universally true law requires that we first know the proportion of observed regularities that are *in fact* universally true laws compared to the proportion of observed regularities that are *not* universally true laws but are rather mere accidental regularities. Otherwise, we could not mount a successful probabilistic argument[48] claiming that observed regularity increases the likelihood of universal truth!

But to know the proportion of observed regularities that are *in fact* universally true laws compared to the proportion of observed regularities that are *not* universally true laws already presupposes we can establish that something is a universally true law – and that is the very thing P sought to give a condition for determining in the first place!

Let's summarize this argument from the meaningfulness of statements of laws of nature against verificationism. For a verificationist to account for the truth and hence meaningfulness of universal laws of nature, he or she has to allow probabilistic observational verification. But to move from probabilistic observational verification and infer the truth of a universal law of nature, the verificationist needs a linking premise, P, that gives us a link between observed regularity and the likely truth that the given observed regularity is a universal law. But in order to justify this linking premise (i.e. the premise which allows us to conclude something is (likely) a law of nature), we already have to have prior knowledge of the proportion of observed regularities which are in fact laws compared to the proportion of regularities which are not. But again, knowing this *itself* presupposes prior knowledge that certain regularities are laws of nature, in which case P itself *presupposes* we have prior knowledge of laws of nature. But P was supposed to give us a condition for inferring the truth of laws of nature *in the first place*! We have vicious circularity: To know P, we need knowledge of laws of nature; but to have knowledge of laws of nature, we need to know P.

So, under verificationism, to infer the truth of any law of nature (and hence to begin to know the truth of any law of nature) requires that we

already possess a prior knowledge of laws of nature in the first place – and that's absurd. Hence, verificationism cannot account for the truth and hence meaningfulness of laws of nature, and since the truth of such laws is essential to science, it follows that verificationism undermines science itself.

Vicious regress

Consider, further, that the verification condition of P itself either makes P probably true or conclusively makes it true. Clearly it cannot be the latter, since P *itself* is a universal statement that applies to any and all observed regularities, and since universal statements (as the verificationist has conceded) cannot be conclusively observationally verified, we must reject this horn of the dilemma.

But now consider the case in which the verification condition of P itself makes P probably true. A vicious regress looms on the horizon. For, yet again, P itself is universal and is purportedly true of a potential infinity of observed regularities. But in order for some set of observations to make P probably true, we would yet again (just like in the case of inferring the universal truth of a law of nature) have to amass a host of particular instances of P's being true and infer that, therefore, probably P is universally true. But this inference, just like the laws of nature case, requires a *further* linking premise, P*, of the sort that *if it is an observed regularity in a finite number of as-yet-experienced instances that P holds true, then, probably, P is universally true*.

But the same dilemma now faces the meaningfulness of P*, since in order to be meaningful it must be observationally verifiable. And, yet again, P* is universally true just like P and the purported universally true law of nature. So the same dilemma will yet again force us to utilize a further linking premise P**, and so on ad infinitum. This shows that verificationism cannot account for the meaningfulness of universally true laws of nature in *yet another sense*, since accounting for their universal truth requires the observational verification of an infinite number of linking premises, which is absurd.

Falsification

Nor will it do to suggest rather that the universal law is meaningful in virtue of the possibility of observationally falsifying it, since this criterion itself faces difficulties. For instance, existential generalizations such as *there is at least one purple polar bear* cannot strictly speaking be observationally falsified. This is because to observationally *falsify* this statement would require us to observationally *verify* that *all polar bears are non-purple*, which again has the same universality and thus non-observational-verifiability that we were trying to avoid in the first place.

A disjunctive solution?

In light of these difficulties, the verificationist may change his or her analysis of meaning to an analysis in terms of a *disjunction*. Roughly, this would modify VP to:

> *VP**: A statement is meaningful provided that *either* some set of observations could establish its truth *or* some set of observations could establish its falsity.

This analysis allows the verificationist to countenance the meaningfulness of both universal statements (since such statements can often be observationally falsified) and existential statements (since such statements can often be observational verified). But there are major difficulties facing such a move.

Consider first that there exist clear counter-examples to the disjunctive view, as there is still a difficulty in accounting for the meaning of sentences that contain both universal *and* existential quantification. Philosopher Jeff Speaks gives the following examples (Speaks (2011)):

> *For every question, there is an answer.*

This can be formalized as follows:

For all x, if x is a question, then there exists a y such that y is an answer to x.

Now, this claim clearly cannot be established to be true by any set of observations, since it is just as much of a universal statement regarding all actual and possible questions. Again, no finite set of observations could establish the truth of such a statement concerning *all* questions. Thus, if this statement is to be meaningful under VP*, it must instead be able to be observationally established false (i.e. conclusively falsifiable). But it cannot be observationally established false. To observationally establish the falsity of a claim is just to observationally establish the truth of its negation from a set of observations. In this case, though, we would have to derive the following:

It is not the case that for all x, if x is a question, then there exists a y such that y is an answer to x.

But this is equivalent to:

There exists at least one x such that it is not the case that if x is a question, then there exists a y such that y is an answer to x.

But this, in turn, is identical to following statement (since an implication is false iff its antecedent is true while its consequent is false:

There exists at least one x such that x is a question and there does not exist a y such that y is an answer to x.

And this, moreover, is equivalent to the following statement. Call the following statement 'S':

S: There exists at least one x such that x is a question and for all y, it is not the case that y is an answer to x.

Therefore, for the original sentence to be able to be observationally established false (i.e. conclusively falsifiable), we must be able to derive S from some set of observations. But we cannot do so, since to derive S from

a set of observations requires us to be able to "derive from such a set the *universal* claim that *there is some question such that every answer fails to be an answer to that question* [this is, in fact, what S states]. But we can no more derive this from a set of observation sentences than we can derive 'All swans are white.' from such a claim" (Speaks (2011)).

The diagnosis of this problem, as Speaks points out, is that to establish the falsity of a sentence is just to establish the truth of its negation; and provided that a statement contains both an existential and universal quantifier (as our example did), then both that statement and its negation will be universal statements. Hence, the disjunctive view of verificationism, which holds that *a statement is meaningful provided that either some set of observations would establish its truth or some set of observations would establish its falsity*, makes such mixed quantification statements meaningless.

To summarize, this is because establishing the *truth* of such mixed quantification statements requires establishing the truth of a universal statement by means of a finite set of observations (which is impossible, as seen earlier); and to establish the *falsity* of such mixed quantification statements also requires establishing the truth of a universal statement by means of a finite set of observations (which, again, is impossible). Therefore, mixed quantification statements are meaningless on this analysis. But that is absurd, since statements like 'there is a question which has no answer' (this is an abbreviated version of S) are clearly meaningful.

Further still…

There is a yet further way in which verificationism cannot account for the laws of nature. Laws of nature support counterfactual conditionals insofar as such laws entail truths such as *had I let go of my water bottle yesterday while standing on the surface of the earth, and had there been no countervailing forces present, my water bottle would have accelerated at 9.8 m/s/s towards the ground*. Clearly, we cannot observationally verify such statements. No past individual, moreover, can verify all counterfactual conditionals (since there's an infinite number of counterfactual conditionals

implicated in a single act), and so even in any act for which we can construct meaningful and true counterfactuals, no one can in principle verify their truth in an act of observation (since, again, the very act of observing a state of affairs entails the non-occurrence and hence non-observability of counterfactual ones).

The verificationist may respond by altering his or her position again, stating that counterfactuals are meaningful in virtue of the fact that, *were a given counterfactual state of affairs to obtain, it would be observationally verifiable*. Call this fact 'F'.

But this response is no good either, since F *itself* is a counterfactual conditional, and hence it cannot *ground* the meaningfulness of counterfactuals since it *presupposes* their meaningfulness in the first place.

Moreover, arguably this response entails a vicious regress, since now we have a counterfactual conditional, F, which itself would only be meaningful (per its own terms) in virtue of a *higher order* counterfactual conditional, i.e. in virtue of the the fact that *were the counterfactual F's antecedent to obtain, then it would be observationally verifiable*. But the specter of a vicious regress looms, since this fact, again, is itself yet another counterfactual and would thus only be meaningful in virtue of a yet higher order counterfactual. And so on ad infinitum.

Finally, if we allow that the verification condition can be counterfactual, then why shouldn't we hold that the verification condition for statements about supernatural beings (say) is the set of observations you would make if you were another supernatural being? Verificationism is threatened with trivialization by allowing counterfactuals into its analysis of meaning.

3.2.3 Verificationism undermines science: unobservable entities

The third and final way that verificationism undermines science is as follows. Scientific theories, as Edward Feser points out, "sometimes postulate unobservable entities. [But] because they are unobservable, a strict interpretation of the principle of verifiability [the verification principle] would seem to render statements about such entities meaningless

if interpreted in a realist way" (Feser (2019)). An example of such unobservable entities would be quarks, which are subatomic particles that comprise neutrons and protons. But, as Feser points out, "there are powerful arguments for realism about unobservable entities, which constitutes a further reason to reject the principle of verifiability" (2019, p. 144).

Other reasons for rejecting verificationism are entailed by the considerations against scientism that will be expounded in later sections. The point to note for now is that verificationism is false, and hence the inference from its truth to scientism's truth is blocked. Note, further, that we have employed in this section (section 3.2) a whole host of techniques explored in earlier chapters for analyzing and evaluating arguments – counter-examples, vicious regresses, parallel arguments, analogies, dichotomies, self-refutation, question-begging, non-sequiturs, formal logic, identification of implicit assumptions, and so on.

3.3 The nature of science: a response

The argument for scientism on the basis of the nature of science roughly proceeds as follows:

1. Science is characterized primarily by objectivity while philosophy and other disciplines are characterized primarily by subjectivity.
2. Whatever is characterized by objectivity counts as a reliable guide to truth while whatever is characterized by subjectivity does not count as a reliable guide to truth.
3. Therefore, science is the only reliable guide to truth.

As stated, though, such an argument clearly lacks force. For consider the following argument:

1. Empirical observation as a purported way of knowing is characterized primarily by objectivity while examining one's own mental state as a purported way of knowing is characterized primarily by subjectivity.

2. Whatever is characterized by objectivity counts as a reliable guide to truth (i.e. a genuine way of knowing) while whatever is characterized by subjectivity does not count as a reliable guide to truth (i.e. not a genuine way of knowing).

3. Therefore, empirical observation counts as a reliable guide to truth (i.e. is a genuine way of knowing) while examining one's own mental state does not count as a reliable guide to truth (i.e. is not a genuine way of knowing).

But this conclusion is clearly absurd, since knowing our own mental state is one of the most *certain* types of knowledge that exists and hence is most definitely a reliable way of generating knowledge (i.e. knowledge concerning one's own thoughts, ideas, beliefs, desires, emotions, feelings, and so on). And since the argument parallels the first, the first is absurd as well.

The underlying point here is that the mere presence of subjective elements within a domain of knowledge or inquiry neither invalidates nor renders unreliable the domain in question. It simply does not follow, then, from the mere fact (if it is one) that science is objective while philosophy and other disciplines are subjective that only science is a reliable guide to truth.

Misconceiving non-scientific disciplines

The second manner in which the argument from the nature of science fails is that it misconceives the nature of philosophy and other disciplines. Philosophy and other disciplines are, in fact, characterized by significant rigor and objectivity. Philosophy not only regularly analyzes and incorporates into itself the findings of science but also employs the reasoning faculties, concepts, intuitions, and universal experiences common to all humans. The backbone of philosophy, moreover, is logic – and logic is patently objective. Overall, then, while philosophy may at times incorporate subjective elements into its methods and arguments (such as phenomenological inspection or appeals to one's internal ethical intuitions), it is simply false to say it is *primarily* characterized by such subjective

elements. It is, rather, primarily characterized by the use of reason, universal human experience, and conceptual analysis – none of which are subjective in a sense that subverts philosophy's reliability.

Science's presuppositions

A third criticism of the argument from the nature of science is that *science itself presupposes the reliability of subjective elements*. If, therefore, a reliance on subjective elements suffices to undermine a discipline, science itself would be undermined. The reason science presupposes the reliability of subjective elements is that empirical observation itself presupposes the private, first-person, subjective mental states of the observing scientists. Thus, if such knowledge is unreliable, then the very evidential foundation of science – observation – is itself unreliable.

Those beholden to scientism may respond that, even though science presupposes subjectivity in the above sense, science – unlike philosophy – is amenable to independent observational verification by other observers. Now, while it is true that independent observational verification raises our confidence in the deliverances of our observational faculties, this just misses the point.

For, consider that the independent observations of other people can only provide evidence of the veridicality of one's own observation if the subjective mental experiences of those people are, indeed, reliable sources of knowledge. If their private mental states were unreliable or not genuine ways of producing knowledge, then their observational testimonies would be equally unreliable, since their observational testimonies presuppose the private experiential character of their observations. Increasing the number of people will not make a difference either, since increasing the number of unreliable testimonies in the hopes of increasing our confidence in the veridicality of a given observation is like increasing the number of defective batteries in the hopes of being able to charge a flashlight. Overall, therefore, if the presence of any subjective elements render the deliverances of a faculty unreliable (per the argument from the nature of science), then the evidential basis of science is itself unreliable since the evidential basis of

science presupposes the reliability of the private mental states, experiences, and thinking processes of scientists.

Tainting effects

But the criticisms of the argument from the nature of science run deeper still. For starters, the argument emphasizes that the methods of science reduce the tainting effects of confounding biases, personal history, and so on, while other disciplines have no such built-in guarantee. But this argument massively overstates the effects that such biases have on philosophy and other disciplines. Philosophy, like science, is a social enterprise. This means that philosophy, no less than science, centers on discussion among philosophers with vastly different backgrounds and beliefs. This discussion (which manifests itself in email correspondences, live discussions, paper exchanges, lectures, conferences, presentations, and so on) in turn allows philosophers to reduce the sociological, personal, and psychological factors that could potentially influence their conclusions.

Philosophy, like science, also has anonymous referees and peer review. Philosophy, then, benefits from many of the same mechanisms that reduce personal bias within science. This directly contravenes one of the claims adduced in favor of this argument for scientism, namely that 'any philosophical position is mired in the subjective and personal preferences, beliefs, intuitions, and history of the individual philosopher who proposed it'. But even supposing this statement to be true, it would be demeaning to assume that professional philosophers are unable to step outside of their own biases at least to a significant extent to evaluate their own positions with rational scrutiny. I would aver that the argument, then, unfairly attributes a level of incompetence to philosophers that is not actually present in reality.

Technical assistance

Turning now to the point that scientific observations are more technically assisted and meticulously procured compared to more mundane experience, it can be seen that this claim no more supports scientism than

the claims adduced above. First, merely from being more carefully procured and technically assisted by instrumentation, it does not follow that any other discipline which lacks such features can give no knowledge about reality. More importantly, though, is the fact that observations gathered by means of scientific instruments like microscope, telescopes, mass spectrometers, and so on *presuppose* the reliability of more mundane forms of experience, since one cannot even read or collect data from any instrument without utilizing his or her unassisted sensory experience. Far from showing the unreliability of mundane, unassisted experience, then, science cannot function *without* its reliability. And since philosophy and other disciplines often employ within their methods and arguments the universal, unassisted experiences common to all humans, this argument fails to establish the unreliability of non-scientific disciplines.

Disagreement

The final consideration leveled in favor of scientism on the basis of the nature of science was the argument from agreement and disagreement. Because science is characterized by massive agreement while non-scientific disciplines like philosophy are characterized by massive disagreement, it is argued that science is the only reliable guide to reality.

The first thing to consider is that there is actually much more disagreement within science and much more agreement within philosophy than this argument lets on. Nevertheless, it does capture the truth that, in general, there seems to be more agreement within science than philosophy and other non-scientific disciplines. What, then, explains this discrepancy?

The argument from agreement and disagreement accounts for this discrepancy by inferring that all non-scientific disciplines are unreliable guides to reality. But to infer the unreliability of a domain of inquiry from the mere fact of disagreement within that domain is a manifest non-sequitur.

In fact, the claim that the mere existence of disagreement in a given area entails that the area is not or cannot produce knowledge entails that certain areas of contemporary and historical science are not or cannot produce

knowledge. For one thing, the cutting edge of research in almost every scientific field, precisely because such research is merely in its infancy, is characterized by profound disagreement. In addition, there are countless examples of long-standing disagreements within science that still persist to this day, such as the competing theories on the pathophysiology of ADHD. If the mere existence of disagreement entailed unreliability, then substantial portions of science itself, along with clear cases of sound scientific inquiry and methodology, would count as unreliable. But by definition scientism states that sound scientific inquiry and methodology are reliable guides to reality. Hence, if the defender of scientism wants to maintain his or her position, he or she must reject the claim that the mere existence of disagreement entails unreliability. But in that case, the argument in question fails since it needs this claim in order to establish the unreliability of philosophy and other non-scientific disciplines.

In fact, this argument is self-refuting, since there is substantial disagreement about the truth of scientism itself. So, if the mere existence of substantial disagreement within a domain of inquiry entails that the domain of inquiry cannot generate knowledge about reality, it follows that the debate concerning whether or not scientism is true cannot generate knowledge about reality. In that case, though, the defender of scientism has just established that his or her very defense of scientism cannot produce knowledge and thus that he or she cannot know scientism to be true.

Indeed, the argument is self-defeating in yet another way, since the claim that only disciplines susceptible of consensus answers are reliable is *itself* not a scientific claim, but rather a philosophical claim in need of philosophical defense.

We can see, therefore, that the mere fact of disagreement within philosophy and other disciplines does not suffice to show their unreliability. The question, however, remains: what accounts for the real discrepancy between philosophy and science in terms of expert agreement? While we cannot avail ourselves of an extended discussion of this question, a few observations are in order.

First, philosophy deals in large part with interior certainties, i.e. truths that we unreflectively know with direct immediacy as internal convictions. An example of such an interior certainty would be the law of non-contradiction. We know with direct immediacy and (at least initially) in an informal, non-verbal way that something cannot be both true and false in the same respect.

Importantly, though, we only know such interior certainties initially in an informal, non-verbal, unreflective, intuitive way. Such truths are not externally sensible and hence can only be conveyed by means of words and language, both of which are somewhat foreign to the certainties themselves. We are not presented the law of non-contradiction, for example, in a ready-made abstract formula; rather, we primarily intuit its truth and applicability in concrete instances.

This tension between interior truths and their abstract and language-bound formalizations manifests itself in significant portions of philosophy. Philosophy's abstract nature, then, combined with its reliance on language, mean that ambiguities, imprecisions, and so on are bound to arise. It is harder to capture our interior certainties by means of abstract language since not only do we feel their force most strongly solely in concrete instances, but also because our internal convictions of them generally do not consist in language-bound principles but are rather apprehended in a single intellectual act – intuitively, immediately, and without inference (Augros (2017)).

The second observation is actually a cluster of points expressed forcefully by philosopher Herman Cappelen in his article "Disagreement in Philosophy: An Optimistic Perspective". First, the actual empirical evidence for widespread and persistent disagreement within philosophy is quite slender. Second, for the persistent disagreement that does exist within philosophy, a plausible explanation (in addition to the one cited above) results from the *failure of evidence neutrality*. In short, many of the theories we accept will affect what we even *count* as evidence for a given hypothesis in the first place. Finally, philosophical convergence has provided the impetus for spawning special sciences. "In the history of philosophy,"

writes Cappelen, "the following has happened many times: philosophers work on a set of problems and then at some point they converge on precise questions and develop some shared methodological standards… [T]hat set of issues becomes a new discipline." Key aspects of contemporary economics, psychology, biology, sociology, linguistics, and even physics were, historically, philosophical questions. It would be "arbitrary, then, to not treat that as an instance of philosophical convergence."

It is clear from these observations, though, that philosophical disagreement is explained by a whole host of factors that do not involve the unreliability of the entire discipline or the ways of knowing contained therein. It is clear from this whole discussion, moreover, that the main arguments in favor of scientism all fail. The question of whether there are successful argument against scientism, however, is yet to be explored. It is to this issue that we shall turn next in order to conclude our case study on scientism.

4 Arguments Against Scientism

4.1 Self-defeat

According to scientism, science is the sole reliable guide to reality. But no result of empirical observation or scientific experiment could possibly establish this claim. No measurements of pressure, temperature, rates of chemical reactions, light transmittance, entropy, galactic red shift, or any other physical quantity could possibly give a result which entails that science is the only source of knowledge. Hence, scientism can only be established by non-scientific means, in which case scientism subverts itself, since (per its own criterion) non-scientific means are unreliable.

Additionally, science cannot possibly establish that science *itself* is a reliable guide to reality, since establishing such a conclusion on the *basis* of science presupposes science's reliability *in the first place*. In other words, it would be circular to prove science's reliability by means of science, since such an endeavor presupposes from the get-go the prior reliability of science. But if science cannot even prove its own reliability, it clearly

cannot prove it is the *only* reliable guide to reality. But in that case, the claim that science is the only reliable guide to reality is not a scientific claim, and hence per scientism's own lights it cannot be reliably known. Scientism is thus self-undermining in yet another way.

Now, those beholden to scientism may insist that any non-scientific discipline implicated in the justification of science's reliability must itself really just be a part of science – we should, in essence, expand or broaden our concept of science to include any discipline implicated in such an endeavor.

The trouble with this proposal is twofold. First, it merely pushes the problem back a step.[49] To see this, label all the non-scientific disciplines implicated in the justification of the reliability of science as 'D'. The proposal in question states that we really should broaden our concept of science to include D. Thus, let's label the combination of science itself (as previously and more narrowly conceived) with D as science*.

The trouble, though, simply re-arises, since science* itself cannot justify its own reliability any more than *science* originally could. For science* to justify its own reliability would presuppose its own reliability in the first place, and thus to justify the reliability of science* requires going beyond science* itself. But now we must appeal to non-scientific* knowledge and reasoning, which is the very thing the proponent of scientism sought to avoid in the proposal of science*.

The proponent of scientism may modify science* to include an even broader conception of science, science**. But the same argument would apply, and a vicious regress clearly looms on the horizon. We must ultimately, therefore, countenance the existence and reliability of non-scientific knowledge.

The second difficulty with this proposal is that it trivializes the thesis of scientism. It is guilty (as it were) of "arbitrarily redefining 'science' so that it includes anything that could be put forward as evidence against scientism" (Feser (2017)). The proponent of scientism is sliding closer to

the position that any sort of rational inquiry is 'science', in which case it is trivial and uninteresting to say that only science provides knowledge of reality. Because rational inquiry is defined as domains of investigation which provide knowledge, to define science as any sort of rational inquiry is just to define science as domains of investigation which provide knowledge. But in that case, the thesis that 'science is the only source of knowledge' is the triviality that 'domains of investigation which are sources of knowledge are the only sources sources of knowledge.'

4.2 The nature of science: abstractions and presuppositions

We saw earlier that the nature of science lends no support to scientism. On the contrary, the nature of science provides a number of cogent considerations *against* scientism. The first aspect of the nature of science (more specifically, physics) that undermines scientism is its methodological limitations. "Physics," writes Feser, "insists upon a purely *quantitative* description of the world, regarding mathematics as the language in which the 'Book of Nature' is written... Hence, it is hardly surprising that physics, more than other disciplines, has discovered those aspects of reality susceptible of the prediction and control characteristic of quantifiable phenomena" (Feser (2017)).

Our experience of the world, by contrast, contains a whole host of qualitative characteristics, as we perceive colors, flavors, odors, sounds, warmth, and so on. To achieve its exact and precise nature, however, physics must abstract "from these rich concrete details, ignoring whatever cannot be expressed in terms of equations and the like and thereby radically simplifying the natural order" (Feser (2017)).

Now, there is nothing wrong with this – in fact, it is one of the triumphs of physics. But we cannot forget that the deliverances of physics are in large part mathematical abstractions and hence do not capture or describe the whole of concrete, physical reality from which it abstracts. "Abstraction," emphasizes Feser, "by its very nature involves ignoring the individualizing features of a thing and focusing instead on patterns it has in common with other things" (Feser (2019)). It is clear, then, that the very nature of physics

guarantees that portions of reality fall through its methodological net (as it were). This means, in turn, that physics (and the argument can be extended to science itself) cannot provide a wholly exhaustive account of nature. This, in turn, gives us reason to reject scientism since scientism holds that science alone exhausts our knowledge of reality.

In addition to the abstractions and simplifications inherent within science, there is a second aspect to the nature of science that subverts scientism. Consider first that science itself rests on a whole host of presuppositions without which it could not function. Science must, for instance, presuppose the existence of an external world, the existence of other minds, the reliability of our reasoning faculties and sensory apparatuses, the reliability of memory and inductive reasoning, the uniformity of nature, and so on.

Take memory, for instance. No scientific experiment or investigation could establish memory's reliability, since gathering and recording data, discursively reasoning premise by premise to scientific conclusions, and so on *presuppose* memory's reliability in the first place. A similar argument extends to the other members on the aforementioned list of science's presuppositions.

But since scientific methodology *presupposes* the truth of these propositions, it cannot *justify* or *establish* their truth on pain of circularity. But if the only reliably known things are those revealed to us by science, it follows that the entire enterprise of science rests on foundational presuppositions which are unreliable, in which case science itself is unreliable. Scientism thus undermines the very practice of science.

Finally, not only does science depend upon philosophy for the justification of its presuppositions, but it also depends upon philosophy to a large extent for the interpretation of its results. The reason science needs philosophy for the proper interpretation for its results is that the findings of science implicate us in metaphysical debates about the nature of material reality, universals and abstracta, scientific realism, and so on. Feser

establishes this dependence by means of a string of questions, and they are worth quoting at length:

> For example, is the world fundamentally comprised of substances or events? What is it to be a 'cause'? What is the nature of the universals referred to in scientific laws—concepts like *quark*, *electron*, *atom*, and so on? Do they exist over and above the particular things that instantiate them? Do scientific theories really give us a description of objective reality in the first place, or are they just useful tools for predicting the course of experience? Scientific findings can shed light on such metaphysical questions, but can never fully answer them (Feser (2017)).

In examining the nature of science, therefore, a few major insights can be gleaned. First, no scientific experiment or observation can establish that science is the only reliable guide to reality. Second, science itself cannot establish its own reliability. Third, science is fraught with presuppositions, none of which can be justified scientifically. Finally, even the proper interpretation of scientific results requires recourse to philosophy. Each of these four insights, in turn, establish the falsity of scientism.

4.3 Contra scientism: the nature of the mind

If we restrict ourselves to *scientific explanations* of how mindless motions could give rise to subjective, first-person selves, then it seems we will, as a matter of conceptual necessity, forever remain in the dark. In the following argument, we shall understand science in the same sense as used throughout this case study (i.e. empirical testing and experimentation grounded in observation). The term 'empirical instruments' is understood as humans' perceptual apparatuses in addition to the tools and machines scientists use like microscopes, telescopes, NMR spectrometers, and so on. The argument, then, proceeds as follows:

1. Scientific investigation, by its nature, is limited to that which is detectable by empirical instruments.

2. Only that which is public, objective, and third-person is detectable by empirical instruments.

3. So, scientific investigation, by its nature, is limited to that which is public, objective, and third-person.

4. But the mind, by its nature, is private, subjective, and first-person.

5. If the mind is private, subjective, and first-person, but scientific investigation is limited to that which is public, objective, and third-person, then the mind (*qua* mind) is beyond the scope of scientific investigation.

6. So, the mind (*qua* mind) is beyond the scope of scientific investigation.

7. Whatever is beyond the scope of scientific investigation cannot be fully and adequately explained solely in terms of science.

8. So, the mind (*qua* mind) cannot be fully and adequately explained solely in terms of science.

The basic idea behind the argument is that the scientific method, by its very nature, cannot access the private, subjective (aspects of) mental states and thus cannot fully illuminate their nature. This is because the scientific method is restricted to the investigation of entities and processes that are detectable by our perceptual capacities and technological instruments. But we can only perceive and detect entities, properties, and processes which are public (as opposed to private), third-person (as opposed to first-person), and objective (as opposed to subjective). This is evinced by the fact that only *you* have privileged access to your mental states – no one else possesses such access. Their outsider's view (as it were) is restricted solely to public, objective facts about you. Because of this, the private character of mental states is, strictly speaking, beyond the scope of scientific investigation, in which case there cannot be a full and adequate explanation of mental states in purely scientific terms.

But, one may object, it is *obvious* that science can investigate the nature of the mind! We know, for instance, that certain neurophysiological impairments lead to corresponding mental impairments. And we know this through the methods and tools of science.

There are a number of things to be said in response. Note, first, that the argument never claimed science is silent on *all* aspects of the mind. Rather, it argued that the mind cannot be *fully* accounted for by science. Thus, the argument is compatible with scientific insights into the nature and activity of the mind.

More importantly, though, none of these correlations are discovered by scientifically testing the mind *qua mind*. They are discovered by destroying or observing the destruction of discrete patches of neurons and looking at corresponding behavioral effects. This is not scientific testing or experimentation of the mind *qua* the mind (as they are not poking the *subjectivity* of mental states with their probes). Rather, this is scientific testing of the mind *qua its manifestations in behavior and neurophysiology*.

Third, even when science seems to investigate the nature of the mind *qua* the mind, this can only be done indirectly through an intermediate which, necessarily, renders such an investigation not a scientific study of the mind *qua mind* (i.e. the subjective mental states themselves). Take, for instance, the striking correlations referenced above. The only reason we know of the correlations between brain damage and mental impairment is because *patients report their inner, subjective, private, first-person states to researchers*. It is *not* scientific experimentation which discovers, solely from testing a mind in itself (from testing a mind *qua* mind in all its irreducible subjectivity) that we know such correlations exist. Rather, it is an inference from the *objective, third-person, public speech reports* that the patients give of their inner lives. This is not, then, a successful objection to the argument.

Note, further, that this argument is actually compatible with physicalism. It just shows we need to go beyond pure empirical methodology and the scientific conceptual apparatus to fully account for (explain, gain knowledge of, etc.) the mind. But given that this is the case, it follows that scientism is false, since our account of reality cannot be restricted to a purely scientific one.

4.3.1 The objection from behaviorism

In light of the above argument, one must either eliminate the private, subjective mental states from his or her account of the world or reject scientism. Now, some proponents of scientism may bite the bullet and eliminate the inner qualitative character of conscious experience, instead opting for an account of the mind known as behaviorism. In fact, the twentieth century verificationists famously adopted behaviorism precisely because their account of meaning as observational verifiability could not account for the private and subjective nature of mental states.

According to behaviorism, mental states *just are* behavior – there is nothing more to them. All there is to the mind, for the behaviorist, is just actual or potential observable behavior. So, when we say that Elijah is happy, all this means is (perhaps) the following patterns of behavior or dispositions toward behavior obtain: Elijah is smiling and laughing; Elijah is acting in a hyperactive and bouncy manner; Elijah's verbal expressions cohere with his being happy; there are changes in his pulse, tone of voice, pupil dilation, etc.; and so on. Behaviorism provides an avenue for the proponent of scientism to resist the argument from the nature of the mind because it holds that minds solely consist in third-person, objective, public facts about people (namely, their behavior) which are amenable to scientific investigation.

A full analysis of behaviorism will have to wait for chapter five. Suffice it to note for now that behaviorism is deeply implausible, and hence scientism won't be able to avoid the argument from the nature of the mind by appealing to it.

4.4 Logic and mathematics

Can you taste a logical principle? For instance, can you taste modus ponens?[50] It seems like a silly question, but it gets to the heart of a central difficulty facing scientism. From the silliness of the question, we can immediately see that modus ponens is simply not the *kind* of thing that can be tasted. Similarly, though, we cannot feel, touch, see, hear, or smell modus ponens (think: can you squeeze modus ponens? What is the size or

shape of modus ponens? What is the weight of modus ponens? What is the smell of modus ponens? Does it smell more like lemons or lilacs?).

These silly questions reveal to us something profound about the nature of logical principles: they are not amenable to empirical investigation, experimentation, and testing. This is because empirical investigation presupposes that the objects under investigation are sensible either directly (by means of our five senses) or indirectly (with the aid of scientific instruments like microscopes, telescopes, and so on). But, as seen by the silliness of the questions posed earlier, modus ponens is neither directly nor indirectly sensible – and the same holds true for every other logical principle.

Logical principles, moreover, are necessary and unchanging, whereas sensible things amenable to empirical observation are contingent and liable to change. It follows, then, that logical principles are not susceptible of empirical observation, in which case they can neither be the objects of scientific study nor knowledge gained as a result of scientific study. But if scientism is true, then the only knowledge is scientific knowledge (i.e. knowledge that is produced by scientific study). Hence, since we have knowledge of logical principles, but yet this knowledge is not produced by scientific study, it follows that scientism is false.

But just as we cannot taste modus ponens, we also cannot taste the number three or the proposition that the derivative of x^2 is 2x. To see this, just ask yourself: What does the derivative of x^2 taste like?

But if we have mathematical knowledge (indeed, the proponent of scientism arguably must countenance mathematical knowledge given that mathematics is essential to scientific investigation), and if mathematics (like logic) is not sensible or empirical, then it follows that scientism is false.

4.5 Ethical and epistemic norms

We normally take ourselves to have ethical knowledge. For instance, most of us suppose that we know that *one ought not torture innocent human beings just for fun*. But just like mathematical and logical principles, we cannot taste, touch, hear, see, or smell ethical principles. Ethical principles like *murder is wrong* are immeasurable and intangible. Like mathematical and logical principles, then, ethical principles are not the objects of scientific investigation, and our knowledge of them is not produced by scientific inquiry either. Thus, there is non-scientific knowledge, in which case scientism is false.

Those who adopt scientism are likely to take one of two routes. They will either deny that we have any ethical knowledge or else hold that science alone can produce ethical knowledge.

Although we shall not explore the former route, suffice it to note for present purposes that it is not only *prima facie* implausible, but it also requires justification – justification which will necessarily implicate the proponent of scientism in philosophical rather than scientific reasoning and inquiry, since no scientific experiment could possibly show that *humans have no moral knowledge*. Arguably, then, the proponent of scientism must take the latter route.

But this route is unacceptable as well. This is because science, by its very nature, is limited in scope to *descriptive* claims about how the world *is*. But many ethical principles are *normative*, i.e. non-descriptive, in that they are claims not about how the world *is* but rather how the world *should* or *ought* to be. Therefore, our knowledge of normative ethical principles must necessarily extend beyond the limits of science.

It is worth dwelling on this point a little further through an example. Science arguably can warrant belief in the following:

1. Humans scream in pain if you hit them with great force.
2. Elijah is a human.

But does it follow from these two that *you ought not hit Elijah*? It doesn't. This would be deriving a normative premise solely from descriptive premises, which is logically invalid. Could we derive it if we reasoned as follows?

1. Humans scream in pain if you hit them with great force.

3. Elijah is a human.
4. Pain is a form of suffering.
5. Elijah really does not want to suffer.

Again, it wouldn't follow. Nothing in those premises alone warrants inferring the normative statement that you *ought not hit Elijah*. In order to make such an inference from (1) and (2) valid, we would have add a premise like *you ought not inflict pain on humans*. But again, since this is a normative statement, no amount of descriptive, scientific statements could entail its truth. In general, any linking premise *from* certain scientific, descriptive facts about pain, suffering, desires, and so on *to* a normative conclusion must *itself* include a normative element, in which case our knowledge of such a linking premise is non-scientific knowledge. Therefore, scientism is false.

Of course, science can *inform* certain normative ethical judgments. For instance, suppose we agree that we should not inflict needless pain on conscious creatures. Call this principle 'P'. Suppose, further, that we find empirical evidence that supports the proposition that animal species S feels pain. In this case, scientific knowledge of S has clearly informed our ethical judgments insofar as it has allowed us to extend the scope of P to include S. The point, however, is that science *alone* cannot determine the truth of P (or, for that matter, any other ethical principle or judgment).

Like ethical norms, epistemic norms also count as non-scientific, and thus the defender of scientism must either deny the existence of epistemically normative knowledge or else admit the falsity of scientism. But unlike ethically normative knowledge, arguably science itself requires epistemically normative knowledge. For instance, arguably science

presupposes normative claims like *scientists ought to report their findings truthfully*. But if this is true, then since such norms are, again, non-scientific, it follows that scientism is false.

4.6 Science, freedom, and moral responsibility

Arguably, knowing that person P is morally responsible requires knowing that P is free. But knowing P is free arguably requires knowing that two conditions obtain. The first is the Alternate Possibilities Condition (APC): P has the capacity to choose or act *otherwise* than P in fact does. The second condition is the Source Condition (SC): P is in conscious *control* of P's actions in the sense that P is the conscious *source* of P's actions, directing them without anything else directly making P act.

Purely scientific investigation, however, could never establish the conjunction of APC and SC for a given person and hence could never give knowledge as to whether the conjunction of APC and SC is met.

Take, for instance, the condition that P has the capacity to act otherwise in a given situation. Precisely because scientific investigation is limited to actual, occurrent events, an experiment could never establish that P has the capacity to act contrary to what P in fact does (as that is non-actual and non-occurrent).

One might respond that, since P's act was wholly unpredictable, it follows that P could have done otherwise. But unpredictability does not entail the capacity to have done otherwise. Weather and environmental conditions are notoriously unpredictable given the seeming infinitude of causal factors contributing to their outcomes; but yet no one would regard weather and environmental systems as free to do otherwise than they in fact do.

One could respond that, while *in practice* we cannot predict weather and environmental outcomes, they are at least *in principle* predictable provided we have a full grasp of all the causal factors influencing the result. P's actions, by contrast, are *in principle* unpredictable.

But this response begs the question at issue, for whether or not P's actions are, in fact, unpredictable in principle is precisely what is at issue in trying to determine whether or not P could have done otherwise.

Moreover, in order to determine whether or not P's actions are unpredictable in principle, one would not only have to identify the myriad causal factors contributing to P's actions, but also calculate how such causal factors specifically interact together to produce an output. One would then have to show that this output ontologically (not merely epistemically) 'leaves open' different actions, thereby making the precise action taken by P unpredictable in principle. And given there are hundreds of billions of neurons with trillions upon trillions of connections, all of which causally impact in some way or other the outcome of a given action (not to mention all the non-neuronal causal factors), the prospects for the completion of this task seem overwhelmingly grim.

But suppose science *were* able to accomplish such a task. There are nevertheless still two major problems for the response in question. First, science cannot establish that there are no causal factors of which current science is either (i) unaware or (ii) cannot detect in principle (say, because they are non-physical). This shows that even if such a monumental task were accomplished, the scientific evidence in and of itself would be insufficient to rule out yet further, unknown external causal factors that determine P's action. Therefore, the scientific investigation *by itself* cannot establish the *in principle unpredictability* of P's action.

The second problem with this response is that *even if,* per impossibile, the scientific investigation could establish that there are no causal factors of which current science is either (i) unaware or (ii) cannot detect in principle, it *still* would not follow that science can establish that P could have *freely* done otherwise. In other words, even unpredictability in principle is insufficient to establish that P could freely have done otherwise. For it could be the case that P's action was objectively random, indeterministic, and chancy. But there is no freedom in random chaos. Even if, say, a die was ontologically chancy and indeterministic, it would not *freely* be able to do otherwise. So, even granting that science could establish P's action was

unpredictable in principle, it still does not follow that P could freely do otherwise. For all this response has shown, P's action could be a chaotic, random process.

Nor will it do to ask P whether or not P could have done otherwise, since whether or not P is correct or mistaken in P's response is precisely what is at issue. So, to merely assume the truth of what P says is not to establish scientifically that P is free but rather merely to assume P is free. Hence, neither unpredictability nor the administration of a questionnaire would suffice to scientifically establish that P has the capacity to do otherwise.

We might think that science can establish P can do otherwise by showing that, in similar situations, P performs different actions; and, moreover, in almost identical situations in the past, P performed different actions. But again, this response merely assumes that every single one of the infinitude of relevant causal factors impinging on P's action are the same in each situation. This assumption is not only unjustified but straightforwardly false given subtle changes in mood, experiences, neurochemistry, time, location, and so on.

Many of the same considerations show that science alone cannot establish that P satisfies condition SC. How could science establish, even in principle, that P's actions are under P's own conscious, self-sourced *control*?

Again, scientists might administer a questionnaire to P or compare P's actions to other cases where P clearly acts involuntarily (and hence acts not in a manner under his or her own control), but both of these are insufficient. In the former case of administering a questionnaire and asking P whether or not P is in control, this merely assumes that P's response is a reliable indicator of whether or not P is in fact in control of P's actions. But this is precisely what is at issue – after all, provided there are a seeming infinitude of causal factors of which P is unaware that affect P's actions, the mere fact that P reports he or she is in control does not entail P is in control.

The latter case of comparing P's actions to other cases where P clearly acts involuntarily is similarly insufficient to establish P acts under his or her own control. For consider what 'voluntary' means in this context. In this context, 'voluntary' just means that P is not aware of any external (i.e. external to the will) thing that causally determines P's actions. But *not being aware of any external thing that causally determines P's actions* does not *by itself* entail *being aware that no external thing causally determines P's actions*, just as *not being aware that any x-rays are passing through you* does not by itself entail *being aware that no x-rays are passing through you*. [51] In other words, science may be able to establish that P is not aware of any external thing determining P to act, and to that extent P acts 'voluntarily'. But this by no means entails that P is, in actuality, in control of his or her own actions provided there is the live possibility in principle that there are causal factors impinging on or determining P's actions of which P is unaware.[52]

Whether or not it is true that P is determined by prior conditions is not what concerns us here. The point, rather, is that science alone is categorically incapable of showing that P's act is voluntary *in the sense* of being causally determined by nothing external to the will (like desires, beliefs, emotions, neurochemical states of the brain, sensations and environmental conditions, the pushes and pulls of particles, and so on).

So, science alone cannot establish that conditions APC and SC are met. But if scientism is true, then the only things that can be known are those which can be established scientifically. So, if scientism is true, then it cannot be known that conditions APC and SC are met. Assume, for reductio[53], that scientism is true. Since knowing P is free requires knowing APC and SC are met[54], and since (per scientism's lights) it cannot be known that APC and SC are true, it follows that we cannot know P is free. But since knowing that P is morally responsible requires knowing that P is free, it follows we cannot know P is morally responsible. What follows is a striking result: If scientism is true, then we cannot know any person is ever morally responsible. But that is surely absurd, for surely you and I both know that, say, Adolf Hitler is morally blameworthy (and hence morally responsible for his actions). Therefore, scientism is false.

One may question whether APC and SC are actually necessary conditions for a conception of freedom that grants moral responsibility. Although adequately addressing this issue would take us too far afield, a few remarks are in order. For starters, even if we deny one of the two as a necessary condition for freedom and moral responsibility, it is still highly plausible that at least one of the two conditions is a necessary condition. So, even if one denies, say, APC as a necessary condition for freedom, it remains highly plausible that SC is still a necessary condition for freedom. But, as I have argued, science alone can establish *neither* of the two conditions, and hence scientism would still entail that we cannot know that any person is ever morally responsible.

Second, there are reasons for taking both conditions to be necessary for freedom and, as a result, moral responsibility. I will briefly sketch a case for the APC condition, but similar reasoning applies to the SC condition as well. Suppose someone or something utterly lacks the capacity to do otherwise. If that is the case, then anything they do is either already determined for them beforehand, or is purely random or chaotic. But if anything one does is already determined for one beforehand, then one cannot justifiably be praised or blamed for one's actions. This is because those actions were inevitable. It would be like blaming someone for the color of their skin, as it amounts to blaming someone for something they had no capacity to change. Similarly, if anything one does is purely random or chaotic, then blaming them would be tantamount to (say) blaming a die for landing on a two. It is simply unwarranted. Both inevitability and random chaos disallow for freedom. By extension, they also disallow for moral responsibility.

4.7 A demonstration of philosophical knowledge

Although the following proof is arguably implicit in a number of things already contained in this case study, it is a valuable and straightforward way to challenge scientism. Let P be the proposition that nothing in philosophy can be proven true. Now, either P is true or P is not true. If P is true, then since P is itself a philosophical proposition (P is a proposition about the nature, scope, and limits of proof, knowledge, and philosophy), and since P

states that no philosophical proposition can be proven true, it follows that P cannot be proven true.

But if P is *not true*, then clearly P cannot be proven true (since something can be *proven true* only if it is, in fact, true!). By constructive dilemma[55], it follows that P cannot be proven true. Now, let Q be the proposition that P cannot be proven true. Since we have proven that P cannot be proven true, we have proven Q. But, yet again, Q is itself a philosophical proposition (as it is a proposition about what can be proven in philosophy and is thereby about the nature, scope, and limits of proof, knowledge, and philosophy). Therefore, we have proven a philosophical proposition (Rasmussen (2019)).

Not only does this demonstrate the existence of philosophical knowledge (thereby refuting the core thesis of scientism), but it also challenges the proponent of scientism. For if we reliably have *some* philosophical knowledge, then what is preventing us in principle from reliably attaining *more* philosophical knowledge in ethics, epistemology, metaphysics, philosophy of religion, and so on? There seems to be no relevant difference between different sorts of philosophical inquiry that would justify a divergent epistemic attitude toward their reliability. But without such a relevant difference, it seems we can reliably have philosophical knowledge in these other philosophical domains. The principle of relevant differences, then, gives us yet further reason to reject scientism.[56]

4.8 Underdetermination and theoretical virtues

The *underdetermination* of theory by evidence is best illustrated by reference to a concrete example. Take, for instance, environmental science and climatology. These disciplines consist in the construction and analysis of climate models and the use of mathematical representations for land and ocean temperature data. Such models and representations often involve idealized and/or simplified simulations which are intended to approximately correspond to reality. These disciplines also make predictions about the future (for example, the extension of global warming trends and rising sea

levels in the coming decades) and generalizations about atmospheric conditions not available for direct, empirical study (for example, conditions on planets well beyond Earth and even our solar system). Finally, these disciplines characteristically utilize inference to the best explanation when they affirm, for instance, that carbon dioxide is the major cause of global warming.

In climatology, though, often there are multiple competing hypotheses, theories, or explanations for which the evidence is, by itself, insufficient to establish which one is true. The competing theories are *underdetermined* by the evidence. How can we know which hypothesis or theory is true when the evidence alone is insufficient to warrant accepting one as opposed to the other?

In cases such as these, there are certain explanatory or theoretical virtues by which we establish which theories are true. As explored in the previous chapter, these include simplicity (the amount, complexity, and nature of claims), predictive and explanatory power, and coherence. Since these theoretical virtues are, necessarily, not part of the empirical evidence base (since they are precisely what we use when the empirical evidence alone does not establish which theory is true), their utilization in producing explanatory or theoretical knowledge is an instance of accepting conclusions which go beyond empirical evidence. This argument, in fact, applies to almost all explanatory and theoretical knowledge, comprising nearly all of the knowledge within the natural sciences.

Relating this to climatology, there may be two hypotheses at play to explain rising temperatures. The first grounds the explanation in terms of the nature and molecular structure of carbon dioxide. The second, by contrast, grounds the explanation in terms of an as-yet undiscovered and, indeed, undetectable type of particle which is attached to carbon dioxide molecules and traps heat *by means of* carbon dioxide molecules. The second hypothesis seems quite clearly false. But why?

Both explanations predict and explain the data (rising temperatures correlated with an increase in carbon dioxide). Neither explanation runs

contrary to climatological knowledge previously gained, and neither is conceptually incoherent. Indeed, any empirical evidence for one will also be empirical evidence for the other.

The answer is that the first hypothesis is much *simpler* than the second. It posits less entities, makes less claims, introduces less complexity, and does not include arbitrary stipulations. In this case, the evidence in and of itself cannot determine which explanation is correct. It is, rather, the explanatory virtue of simplicity that allows us to produce knowledge concerning (in this case) the cause of climate change.[57] Simplicity is, of necessity, not part of the evidence base, since it is a consideration adduced when different explanations or theories are utilized *to make sense of* the very evidence base in question. The same holds true not only for climatology, but for any discipline or area of knowledge in which theory-construction is a central avenue by which knowledge is produced.

Crucially, though, the underdetermination of theory by evidence poses a problem for scientism. Scientific testing, experimentation, prediction, and description are, by themselves, insufficient to establish (at least in many cases) which hypothesis, theory, or explanation is true amongst its competitors. In constructing and accepting scientific theories, then, we must employ theoretical virtues which, by their nature, go *beyond* the scientific observations and experiments. But in doing so, we have now implicated ourselves in a kind of philosophical, not scientific, reasoning, in which case there is non-scientific knowledge (i.e. knowledge that is not justified solely by empirical observation and experimentation). But if there is non-scientific knowledge, then scientism is false. So, scientism is false.

4.9 Counterfactuals

Scientism, like its cousin verificationism, faces difficulties in accounting for the truth of counterfactual statements. No scientific observation or experiment, for example, can establish that *five minutes ago, had I dropped my water bottle, it would have fallen to the ground*. The proponent of scientism may respond in a similar fashion to the verificationist's responses to the problem of counterfactual statements,

however arguably similar rejoinders will equally plague scientism. And if that is the case, then our knowledge of counterfactual statements establishes the falsity of scientism.

4.10 Summing up

In our application of the methods, tools, and skills of critical thinking to scientism, we have come across a kaleidoscopic array of arguments. First, we discovered that scientism is self-defeating, as the thesis of scientism concerns the nature, scope, and limits of knowledge, none of which could be established by scientific investigation alone. We found also that science, by its very nature, is rife with presuppositions and abstractions, both of which necessitate the reality of non-scientific knowledge. We then turned our eyes inward to our own conscious experience. This phenomenological investigation revealed to us that the mind is subjective, first-person, and private. When juxtaposed with an analysis of the limitations of scientific investigation, we reasoned that the nature of the mind compels us to admit the reality of non-scientific knowledge.

Next, we found that even the very *foundations* of science establish the falsity of scientism, since both mathematics and logic are sources of non-scientific knowledge. But just as the very foundations of science prompted us to reject scientism, the normative dimension of reality (which is the foundation of ethics) likewise poses a difficulty for scientism. We then discovered not only that our knowledge of moral responsibility is incompatible with scientism, but also that the principle of relevant differences contravenes scientism.

Finally, we finished this adventure with a consideration of counterfactuals and the underdetermination of theory by evidence, and we saw that to resolve this latter difficulty we had to adduce theoretical virtues (which, again, implicated us in non-scientific reasoning and patterns of argumentation). It is my hope that these arguments not only provide you with an opportunity to practice the critical thinking skills outlined in previous chapters, but also that they improve your vision of the nature of reality and the treasures contained therein.

5 Conclusion

Science, as I stated at the outset of this case study, is beautiful. But as the foregoing discussion indicates, it is not the exclusive discipline that captures the beauties of existence. Non-scientific areas of knowledge are sources of immense beauty as well – indeed, such sources are presupposed by science itself.

In this case study, we began by expounding the predominant motivations behind scientism. These stem from (i) science's predictive, explanatory, and technological success, (ii) the verificationist account of meaning, and finally (iii) science's objectivity and ability to foster expert agreement. Through implementing the philosophical methods and tools explored in previous chapters, we reasoned that none of the above considerations warrant the acceptance of scientism. We then turned to an exploration of the various arguments one may level against scientism, finding that there are indeed strong considerations against it.

It is worth pausing and reflecting on the vast range of methods and tools we utilized in our examination of scientism that were introduced in previous chapters. Among countless others, we drew upon:

- Begging the question
- Non-sequiturs
- Presupposing the very thing in need of explanation
- Soundness, validity, and formal logic
- Explanatory and justificatory circularity
- Counter-examples
- Reductio (entailing a contradiction)
- Construction of parallel arguments
- Symmetry breakers
- Vicious regress
- Self-refutation
- Relevant differences

The value of reflecting on our use of these tools cannot be understated. They are not mere tactics or distinctions that we employ for occasional mental exercise. They have *real application* to serious problems and profound questions concerning knowledge, reality, science, the human person, and more. The majesty of reason, then, is not some abstract or inapplicable pursuit; it guides, shapes, and underpins our worldviews, decisions, and beliefs. It can uncover beauties that would otherwise remain undetectable and treasures that would otherwise remain buried from sight.

Next we turn to a Bayesian problem for a Humean understanding of laws of nature.

Chapter 4: Laws of Nature

1 Introduction

> *"Mathematics possesses not only truth, but supreme beauty – a beauty cold and austere ... sublimely pure and capable of a stern perfection such as only the greatest art can show."* Bertrand Russell, *A History of Western Philosophy*

The correspondence between the heavenly realm of pure and abstract mathematics and the sublunary realm of concrete, physical reality is breathtaking. The elegance and simplicity of mathematical descriptions of physical reality is not only cause for childlike delight (as Bertrand Russell so gracefully captures), but it also underpins the very natural laws governing physical reality. The world seems, as it were, mathematical through and through. But what are these mathematical descriptions represented in science textbooks and found ubiquitously in natural reality? Are such laws of nature merely abstract objects that somehow impose their unconscious will on material reality, governing it from above (as it were)? Or are they instead nothing over and above the mere regularities present in reality? Or perhaps the laws are grounded in the very natures of the objects they describe, detailing their causal powers and characteristic operations that flow from their natures? Do laws of nature even exist?

In exploring these questions, we can come to a deeper and fuller understanding of reality itself. It is through this understanding, in turn, that we heighten our appreciation of the supreme beauty of both the laws of nature and the mathematics in terms of which they are written. "Understanding," as Carl Sagan rightly emphasized, "is a kind of ecstasy," and it is *understanding* that we are principally seeking in this case study.

But before we can delve into one of these questions (the one about mere regularities), we must first briefly touch on a foundational question to any

inquiry: why should we care? Apart from a greater understanding of the nature of reality that can be gleaned from an investigation into the metaphysical status of laws of nature, there are a number of considerations relevant to answering this question. First, scientific laws play a central role in contemporary scientific practice. From the four fundamental forces of nature to the basic constituents of material objects themselves, scientific laws have application universally. Second, laws of nature are key to many explanations of natural phenomena in the world, and thus by gaining a greater understanding of the laws of nature, we thereby gain a greater understanding of the nature of explanation itself and the world in which we live. Finally, one's account of laws impacts other domains of inquiry, such as questions concerning the reliability of induction, the status of counterfactual statements, the existence of free will, and one's broader philosophical worldview. Having covered the importance of this investigation, we can now turn to a few clarificatory notes.

It is worth noting, first, that while many laws of nature are mathematical, by no means does this apply to all laws. *Copper conducts electricity* and *no objects accelerate beyond the speed of light* are examples of laws that are non-mathematical in nature (though, indeed, they are deeply interconnected with mathematical laws describing, for example, electron movement and acceleration respectively). A second thing to emphasize is the *empirical* character of the laws of nature. Such laws are factual, *a posteriori* descriptions of the empirical, concrete world. This contrasts with *a priori* logical truths like *there are infinitely many primes* or *if Smith is taller than Jones, then Jones is shorter than Smith*.

We must recognize, further, that it is one thing to identify a cluster of properties associated with natural laws but another thing to capture the very nature or essence of the laws of nature. Just as a geometer would not capture the very essence of triangularity if he or she told you that triangles come in the varieties scalene, right, and obtuse, that various mathematical proofs have increased our knowledge of triangles, and so on, it likewise will not capture the very essence of natural laws to identify their various properties (like making natural phenomena intelligible, being inferred from observed regularities, being stated in mathematical terms, being universally

quantified, supporting counterfactuals, and providing a sound basis for predictions of natural phenomena). While such properties are interesting and important (indeed, a successful theory of laws of nature must satisfactorily account for each of these properties), we are instead on the search for the very essence and ontological status of the laws of nature. Finally, it is worth keeping in mind the distinction between scientific laws, which are our best approximations to the laws of nature, and the laws of nature themselves, which are the actual features of reality that describe, consist in, or otherwise govern the operations of physical bodies.

There are a number of different ways to carve up the conceptual territory concerning the various approaches to the laws of nature. Nevertheless, we will only explore one argument against a Humean view of laws of nature. Defended and expounded most forcefully by followers of David Hume, the *regularity theory* states that the laws of nature are simply regularities, patterns, or uniformities in the world.

While we won't be exploring any other approaches, it will be useful to understand the alternatives on offer. One approach is the *necessitarian account*, which affirms that natural laws are underlying principles that govern the events and processes in the world and are such as to necessitate the behavior of such events and processes. This approach, in turn, is subdivided into a *platonic account* and a *theological account*. The former holds that laws are constituted by necessary connections between the properties of things understood as abstract entities (i.e. universals or something like platonic forms), whereas the latter views laws as a kind of divine command (imposed from above, as it were, on the events and processes of the world).

Another approach contends that laws are merely useful fictions that neither exist nor are grounded in some underlying explanatory feature(s) of reality. Finally, one other approach is the *Aristotelian account* which sees laws as descriptions of the causal powers that a material object or system tends to manifest given its nature or essence.

In the following sections, we will critically analyze (one version of) the regularity theory. The aim is not only to obtain a greater understanding of the metaphysical questions surrounding the laws of nature, but also to practice the methods, tools, and skills of critical thinking we have been cultivating throughout this journey. In particular, we will pay special attention to Bayes' Theorem.

2 Regularity theory

2.1 Description

The regularity theory that we will examine states that laws of nature are simply regular patterns we happen to find in nature.[58] Nothing causes or grounds them; they're just *there* in nature, and that's that. A statement of a law of nature just describes the way the world is with respect to a uniformity or regularity. A law, for the Humean regularity theorist, is nothing more than a collection of all its instances.

2.2 A Humean motivation

For Hume, a proposition can be justified only by empirical experience or logical deduction.[59] But the claim that there is a necessary connection between two events (objects, processes, etc.) – such that one of them necessitates or makes the other occur – can be justified neither by empirical experience nor logical deduction. Hence, claims to the effect that there are necessary, law-like connections between events (such as all events of copper being exposed to an electric current result in copper conducting electricity) cannot be justified.

Suppose one billiard ball B_1 moves toward another billiard ball B_2 and collides with it. Now, the following isn't logically contradictory: B_1 moves toward B_2, collides with it, but B_2 does *not* move away from B_1. Although all our past experiences with billiard balls tells us that they move away upon collision, none of this experience logically entails that B_2 on this occasion will move away. So, logical deduction alone won't justify a

necessitating, law-like connection between B_1's movement and collision (on the one hand) and B_2's movement (on the other).

But similarly, we can't experience a necessary connection between events. All we ever perceive is the *constant conjunction* of billiard balls moving away (one event) upon collision with another billiard ball (another event). We cannot hear, see, smell, taste, or feel the necessity – all we can observe are events occurring in sequence.

Applying this to the laws of nature, for the necessitarian view, laws involve necessity. If it is a law that all F's are G, then F-hood necessitates G-hood. The Humean argument aims to show that such necessity can never be justified.

Instead of evaluating this argument directly, I'll simply offer a Bayesian argument against this Humean account of laws of nature.

2.3 The problem of persistence

Consider a coin and the following hypothesis H_1: 'this coin is fair, but it will land heads ten times in a row.' This asserts a mere regularity – persistence of the coin in landing heads. Consider a second hypothesis H_2: 'this coin is biased towards heads (its physical constitution is weighted, say, or it has heads on both sides), and it will land heads ten times in a row'.

Suppose I flip the coin and it lands heads nine times. Even after flipping it nine times, the probability of getting heads on the tenth toss conditional upon H_1 is still 0.5. Even after nine consecutive heads, we have no convincing reason to believe that the coin will land heads on the tenth toss under the supposition that the coin is fair. That's pretty straightforward.

Under H_2, however, the probability that the tenth toss will be heads is *greater* than 0.5 (indeed, if the bias in question is specified as heads on both sides, then the probability is 1 (i.e. 100%)).

From the simple illustration above, notice that a greater expectation of heads on the tenth toss is mediated only through a hypothesis that relates the physical constitution of the coin to the frequency of heads. The fact that the coin persistently landed heads doesn't change the *conditional* expectation of the tenth toss landing heads on the hypotheses in question. Even if it's true that nine heads successively come up, it nevertheless remains true that the conditional probability that the tenth toss is heads given H_1 is 0.5. The mere regularity doesn't alter conditional expectations.

But now consider this thought applied in the context of laws of nature. Under the hypothesis that there is no intrinsic or extrinsic reason, explanation, or grounding of a lawful regularity, we simply have no reason to expect the next unexamined instance to abide by the regularity. And this is true even if the regularity has persisted for billions upon billions of instances up until now. Remember, regularities don't alter conditional expectations. The reason is because conditional expectations are a function of what the *hypothesis itself* leads us to expect about reality; it is therefore determined *prior* to examining reality's data.

Under that same hypothesis – the hypothesis that there is no ground or explanation of lawful regularity – we likewise have no reason to expect the next unexamined instance *not* to abide by the regularity. Given this hypothesis, our epistemic position is just like that of the unbiased coin hypothesis: conditional upon this hypothesis, we have no reason or ground for thinking that the tenth, unexamined toss will conform to the regularity – regardless of how strong the regularity is. Such an expectation is justified only provided we have some reason or ground, either intrinsic to the coin itself or extrinsically imposed on the coin, to suppose the coin will conform to the regularity.

Here is the important thing to see: such a reason or ground is precisely what the Humean regularity theory denies. What arises, then, is a problem for the Humean view. We can call it *the problem of persistence*. It starts with a simple question:

Why do things persist or continue operating in accordance with regular laws of nature? Why don't things all of a sudden stop abiding by laws of nature ?

In this context, 'persist' (and its cognates) means continual operation or behavior in accordance with lawful regularities or uniformities. Now, let's get the hypotheses on the table:

> *Humean Regularity Theory (HRT)*: There is no explanation or ground for the regularities in question; nothing *makes it be the case* (i.e. necessitates) that things abide by the lawful regularities.

> ~HRT: There *is* an explanation or ground (either intrinsic or extrinsic to the objects/events in question); something *makes it be the case* (i.e. necessitates) that things abide by the lawful regularities.

Under HRT, each thing genuinely *can* fail to persist in accordance with a given lawful regularity; there is no explanation or grounding that demands something to persist one way as opposed to the other. According to ~HRT, by contrast, things *cannot* fail to persist in accordance with a given lawful regularity.

The basic idea behind the problem of persistence, then, is that if each thing (event, object, process, etc.) can fail to persist in accordance with a regularity, then regularities would have failed to be regularities by now, since it is unlikely for things to continually persist – on and on – in accordance with a given regularity. This is precisely because they genuinely *could fail to do so* at any moment. If there is simply no underlying explanation or grounding of their persistence in accordance with a lawful regularity, then it is *unlikely* for regularities to continue on and on *qua* regularities.[60] If persistence is a brute fact, we have a deep puzzle: how is it that things not only are *able* to persist in accordance with lawful regularities, but also that they *do in fact continue acting/operating in accordance with the lawful regularities*?

But maybe things just persist, and that's that – they just continue operating in accordance with a given regularity, and that's all there is to it.

On closer inspection, though, a probability problem arises for this proposal. This problem requires no assumptions at the outset about whether things can or cannot persist on their own. In fact, this problem is built on the principle of neutrality, where we leave open anything for which we have no evidence for or against.

Consider the following piece of evidence:

Physical objects have persisted in accordance with regular, law-like principles (such as gravitationally attracting one another by an inverse square law; electrons and protons being attracted and repulsed according to uniform patterns; copper uniformly conducting electricity; and so on) for at least one thousand years.

This piece of data seems uncontroversial; even Humeans admit that, thus far, regularities have upheld for quite some time. Gravity has continually operated in accordance with the law-like principles of gravitational attraction for at least a thousand years. Copper has continually behaved so as to conduct electricity; when balls knock into one other on a flat, low friction surface, the impacted ball has always moved away; protons and electrons continue to attract one another and bond in regular ways according to uniform regularities; and so on.[61]

But we are not yet done specifying the relevant data. Consider that persistence in accordance with lawful regularities is not something that only occurs year by year; things also persist at *each and every moment of their existence* by such lawful, physical regularities. The laws of chemistry continue to be followed minute after minute, second after second. This is strange if there is no explanation of persistence, since that means we have lots of moments where the persistence genuinely *can* be broken but *in fact isn't*. In order to capture this insight about persistence moment after moment, I'll convert our data (persistence for a thousand years) into seconds (to make matters simple). So, here is our evidence:

> *Persistence*: Physical objects have persisted in accordance with regular, law-like principles for at least 31,557,600,000 seconds.

Now, if there is no explanation of persistence in accordance with lawful regularities (per HRT), then the probability that things continue to persist in accordance with lawful regularities second after second gets smaller and smaller as the number of seconds increases. Without any grounding or explanation for persistence in accordance with lawful regularities – that is to say, on the supposition of HRT's truth – we have no expectation *one way or the other* as to whether or not physical objects will continue to persist in accordance with the lawful regularities moment after moment, for any moment at which they exist. It's completely open – 50/50. Thus we have no principled way to assign anything *but* a 0.5 probability of persistence in accordance with lawful regularities, one moment after another, given brute, unexplained persistence (HRT).

Indeed, this follows straightforwardly from the principle of indifference[62]. Since HRT affirms that there are no reasons or grounds for expecting persistence, the P(for any second s, physical objects persist in accordance with lawful regularities at s | HRT) is 0.5. And since, according to *Persistence*, there have been 31,557,600,000 such seconds, and since such seconds are probabilistically independent[63], it follows that the P(*Persistence* | HRT) is $0.5^{31,557,600,000}$. That's a *staggeringly small number*.

Again, think of the coin. Under the fair coin hypothesis, each toss could come up heads or tails *as far as we know*. Yet the more it persistently comes up only heads, the more reason we have to think that there is some reason for this, i.e. there is something intrinsic or extrinsic to the coin that predisposes it to persist with heads. The same is so for persistence in accordance with lawful regularities. The more that natural things persist operating in accordance with lawful regularities, the more reason we have to think there is some reason for this, i.e. there is something intrinsic or extrinsic to natural things that predisposes them to persist operating in accordance with such lawful regularities. And that's a problem for HRT, as we've seen.

Consider next ~HRT. Importantly, P(*Persistence* | ~HRT) is one (or nearly one).[64] ~HRT is like the hypothesis that the coin is heads on both sides: it is simply guaranteed that, on each toss, heads will result. Even if we flip the biased (two-sided) coin 8×10^{60} times, the probability that all of them uniformly land heads is one. Similarly, ~HRT *guarantees* that physical objects persist in accordance with lawful regularities at each s, meaning that the probability of *Persistence* given ~HRT is $1^{31,557,600,000}$ – which is just 1.

With this background, we can now use the odds form of Bayes' Theorem to compare HRT and ~HRT. Here is the equation we need to use:

$$\frac{P(\text{HRT} \mid Persistence)}{P(\text{\~HRT} \mid Persistence)} = \frac{P(Persistence \mid \text{HRT})}{P(Persistence \mid \text{\~HRT})} \times \frac{P(\text{HRT})}{P(\text{\~HRT})}$$

Now, let's suppose that the prior probabilities of HRT and ~HRT are approximately equivalent. After all, the two hypothesis *exactly parallel* one another; ~HRT is just the negation of HRT (and vice versa).[65] In that case, the equation now becomes:

$$\frac{P(\text{HRT} \mid Persistence)}{P(\text{\~HRT} \mid Persistence)} = \frac{P(Persistence \mid \text{HRT})}{P(Persistence \mid \text{\~HRT})}$$

And, plugging in the values we obtained earlier, we get:

$$\frac{P(\text{HRT} \mid Persistence)}{P(\text{\~HRT} \mid Persistence)} = \frac{0.5^{31,557,600,000}}{1}$$

After performing the basic calculations outlined in strategy #20, we get that the posterior probability that HRT is true is less than $(6.83 \cdot 10^{-9,499,784,190})\,\%$. That's a decimal followed by almost *9.5 billion* zeroes. Wow.

3 Conclusion

In this chapter, we first canvassed the conceptual landscape in debates concerning the metaphysical status of laws of nature. After that, we looked at one particular view – Humean Regularity Theory. We then found that it

faces a massive probability problem, *the problem of persistence*. It stems from the unlikelihood of continued, brute persistence in accordance with lawful regularities given HRT.

The value of this argument is fourfold. First, it's based on rigorous, contemporary probability theory. Second, it follows a scientific model of empirical confirmation. Third, it is underappreciated – not many people are aware of the power and depth of insights it brings. It is a treasure for truth seekers. Finally, "it can help you develop a sound worldview using the light of reason" (Rasmussen (2017)).[66]

Chapter 5: Mind

1 Introduction

Can you make consciousness out of sand? You can squeeze it, punch it, toss it in the air, or do all sorts of things to it. But is any of that sufficient for conscious experience?

Is mere sand (or other configurations of particles) sufficient for conscious experience? What is the relation between particles and people?

It is to these questions that we turn in this chapter. Before proceeding, it is useful to clarify precisely what we mean when we speak of the mind. By the 'mind' and 'mental phenomena,' I shall understand aspects of the human person such as:

- Intentionality: the directedness or aboutness of thought toward some object
- Qualia: the raw, qualitative, phenomenal 'feel' associated with experiences and other conscious mental states
- Privacy: first-person, privileged, subjective access to one's own conscious states
- Rationality: the ability (i) to intellectually grasp abstract and universal concepts and propositions, and (ii) to apply formal rules of inference

The structure of this chapter is as follows. First, I will critically analyze four theories about the nature of the mind: Descartes' substance dualism, Armstrong's identity theory, Ryle's behaviorism, and Churchland's eliminative materialism. Second, I will consider a potentially groundbreaking argument for the existence of other minds.

2 Descartes' substance dualism

Descartes conceived of the mind (soul, intellect, understanding, *res cogitans*) and material things as separate, distinct substances. The former is an essentially thinking, conceiving, reasoning, spatially unextended, and experiencing thing. The latter, by contrast, is essentially unthinking, unfeeling, spatially extended, and subject to mechanistic laws.

In section 2.1, I will expound upon Descartes' central arguments for dualism (presented primarily in *Meditation II*), whereas in section 2.2 I will critically evaluate them. Finally, in section 2.3, I will evaluate the predominant objections to substance dualism.

2.1 Descartes' central arguments

Descartes provides three central arguments for the non-identity of the mind and the material body. In this section, I outline and motivate these three arguments from (i) conceivability, (ii) dubitability, and (iii) divisibility. The interested reader can consult Descartes' *Meditation II* for further explication.

2.1.1 From conceivability

Descartes' argument from conceivability runs as follows:

1. It is conceivable that my mind exists in the absence of anything material.
2. Whatever is conceivable is possible.
3. So, it is possible that my mind exists in the absence of anything material.
4. The possibility that my mind exists in the absence of anything material entails that my mind is not identical to anything material.
5. So, my mind is not identical to anything material.
6. If my mind is not identical to anything material, then my mind is an immaterial substance and substance dualism is true.
7. My mind is an immaterial substance and substance dualism is true.

2.1.2 From dubitability

Descartes' argument from dubitability runs as follows:

1. My mind does not have the property of being dubitable by me, whereas any material thing has the property of being dubitable by me.
2. If my mind does not have the property of being dubitable by me, whereas any material thing has the property of being dubitable by me, then my mind and any material thing do not share all of their properties in common.
3. Necessarily, if x and y are identical, x and y share all of their properties in common.
4. So, my mind is not identical to any material thing.
5. If my mind is not identical to any material thing, then my mind is an immaterial substance and substance dualism is true.
6. So, my mind is an immaterial substance and substance dualism is true.

Descartes held that his own existence was certainly and indubitably known. For he could not doubt his own existence, since the very act of doubting presupposes his existence as a doubting, thinking thing. On the other hand, he *could* doubt the existence of any material thing, since they could all be the products of a vivid dream or the concoctions of a malicious demon. Hence, material things are dubitable by Descartes, whereas Descartes (qua thinking thing) is not dubitable by Descartes. Hence, Descartes and material things do not share all their properties in common. Thus, reasons Descartes, he must not be identical to anything material (i.e. he is an immaterial mind).

2.1.3 From divisibility

In the argument to follow, x is *divisible* provided that x can be broken down into distinct, self-subsistent, component parts. Descartes argues that mind and material body cannot be identical, as "body is by nature always divisible, and the mind is entirely indivisible. For, as a matter of fact, when I consider the mind, that is to say, myself inasmuch as I am only a thinking

thing, I cannot distinguish in myself any parts, but apprehend myself to be clearly one and entire" (Meditation II). This argument from divisibility proceeds as follows:

1. Anything material is divisible (i.e. can be broken down into distinct, self-subsistent, component material parts).
2. The mind is not divisible (i.e. the mind is an essentially simple and unified thing, not possibly broken down into self-subsistent parts).
3. If the mind is not divisible and anything material is divisible, then my mind and any material thing do not share all of their properties in common.
4. So, my mind and any material thing do not share all of their properties in common.
5. Necessarily, if x and y are identical, x and y share all of their properties in common.
6. So, my mind is not identical to any material thing.
7. If my mind is not identical to any material thing, then my mind is an immaterial substance and substance dualism is true.
8. So, my mind is an immaterial substance and substance dualism is true.

2.2 Critical appraisal of Descartes' arguments

In this section, I will critically evaluate Decsartes' three central arguments. In the sections that follow, each separate criticism will be denoted in italics and numbered.

2.2.1 Conceivability: evaluation

Criticism One: A parody argument

Decsartes' argument is subject to a parody. For starters, note that Decsartes' argument relies on the following thesis, T:

T: If it is possible that x exists in the absence of y, then it cannot be the case that x and y are identical.

For if it *could* be the case that x and y are identical despite the possibility that x exists without y, then premise (4) (which states that x's possible existence without y entails that x and y are not identical) is false. So, Descartes' argument requires T.

But T is logically equivalent to:

If it is possible that x exists in the absence of y, then it is necessarily false that x and y are identical.

By contraposition, this is equivalent to:

If it is not necessarily false that x and y are identical, then it is not possible that x exists in the absence of y.

But since 'it is possible that' is logically equivalent to 'it is not necessarily false that', this is equivalent to:

If it is possible that x and y are identical, then it is not possible that x exists in the absence of y.

Using this logical equivalent of T as premise (4) in a parody argument, we can reason:

1. It is conceivable that my mind is identical to my material brain.
2. Whatever is conceivable is possible.
3. So, it is possible that my mind is identical to my material brain.
4. If it is possible that x and y are identical, then it is not possible that x exists in the absence of y.
5. So, it is not possible that my mind exists in the absence of my material brain.

However, the conjunction of (3) in Descartes' conceivability argument and (5) in this parody argument form a contradiction:

It is possible that my mind exists in the absence of anything material, and it is not possible that my mind exists in the absence of my material brain.

So, the following four theses form an inconsistent tetrad

(1) If it is possible that x exists in the absence of y, then it cannot be the case that x and y are identical. (T)
(2) It is conceivable that the mind is identical to the material brain.
(3) It is conceivable that my mind exists in the absence of anything material.
(4) Whatever is conceivable is possible.

Thesis (1) is deeply plausible and follows from Leibniz' law (since if it's possible that x exists in the absence of y, then x has a property that y does not have, namely the property of existing in a possible world in which y does not exist). Additionally, the falsity of (1) would entail the unsoundness of Descartes' argument. Moreover, Descartes has no principled, non-question-begging way to deny (2).[67]

And the falsity of (3) or (4) would render Descartes' argument unsound. Hence, Descartes is either question-begging or his argument is unsound. Either way, Descartes' argument from conceivability fails.

Criticism Two: Conceivability does not entail possibility

Conceivability does not and cannot entail possibility. There are two reasons for this.

First, our ability to conceive depends on the cluster of descriptions we associate with certain intentional objects. However, unbeknownst to us, two intentional objects which are associated in our minds with different clusters of descriptions can have the same extension in extramental reality. This is the case for intensional contexts in which co-referring expressions (else: co-

extensional[68] intentional objects) cannot be interchanged without potentially changing the truth value of the proposition(s) in question.

With this in mind, consider yet another parody argument that someone would have thought perfectly sensible prior to our knowledge that the morning star is identical to the evening star (Venus):

1. Conceivability entails possibility (assumption for reductio).
2. It is conceivable that the morning star exists in the absence of the evening star.
3. So, it is possible that the morning star exists in the absence of the evening star.
4. If it is possible that the morning star exists in the absence of the evening star, then the morning star and the evening star are not identical.
5. So, the morning star and the evening star are not identical.
6. But the morning star and the evening star *are* identical.
7. So, the morning star and the evening star *are* and *are not* identical, which is absurd.
8. So, conceivability does not entail possibility.

A second argument (leveled below) employs developments in twentieth century modal logic. In particular, the second argument employs the S5 axiom of modal logic that whatever is possibly necessary is necessary *simpliciter*.[69] The second thesis that this argument employs is that identity is essential. In other words, if x and y (rigidly designated) are identical, then it is necessary that x and y are identical. The identity (and identity conditions) of x is an essential and hence necessary feature of x. With these two theses, we can reason as follows:

1. Conceivability entails possibility (assumption for reductio).
2. It is conceivable that the mind is identical to something immaterial (e.g. a res cogitans).

3. It is conceivable that the mind is identical to something material (e.g. the brain).
4. So, possibly, the mind is identical to something immaterial, and possibly, the mind is identical to something material.
5. If x is identical to y, then it is necessary that x is identical to y.
6. So, possibly, it is necessary that the mind is identical to something immaterial, and possibly, it is necessary that the mind is identical to something material.
7. Whatever is possibly necessary is necessary simpliciter. (S5 axiom)
8. Whatever is necessary is actually the case.
9. So, it is actually the case that the mind is identical to something immaterial, and it is actually the case that the mind is identical to something material, which is absurd.
10. So, conceivability does not entail possibility.

Hence, conceivability does not entail possibility, in which case Descartes' premise (2) is false – in which case Descartes' argument is unsound.

Criticism Three: Cannot establish substance dualism

Descartes' arguments require the following thesis in order to establish substance dualism:

If my mind is not identical to anything material, then my mind is an immaterial substance and substance dualism is true.

But this is implausible. For my mind could be distinct from anything material but yet still not be a substance in its own right. For instance, the mind could be an immaterial property (per property dualism), or neutral monism could be true (in which case the immaterial mind could be a non-physical causal aspect of the neutral monist substance), or hylomorphic dualism could be true (in which case the immaterial mind is the human person's substantial form or an operation of the human person's substantial form). Hence, Descartes' premise (6) is highly implausible. Moreover, this criticism applies to all three of Descartes' arguments.

2.2.2 Dubitability: evaluation

Criticism One: A dilemma

As we saw in an earlier chapter, in intensional contexts, co-referring expressions cannot be substituted for one another without potentially changing a proposition's truth value. The failure of intersubstitutivity in intensional contexts derives from the fact that a single object can be presented to a thinking subject under different clusters of descriptions or different aspects. Under such conditions, the agent in question may be unaware that the extramental *referent* of such clusters of descriptions is the same in each case.

In summary, an intentional object (else: linguistic expression) construed *extensionally* is the intentional object's extramental *referent*, whereas an intentional object construed *intensionally* is the mind-dependent cluster of descriptions that the thinking subject *associates* with an object.

With the conceptual groundwork laid, a dilemma now plagues Descartes' argument. The intentional objects of Descartes' doubt (the mind and anything material, respectively) are either construed intensionally or extensionally.

If they are construed extensionally, then Descartes' premise (1) is patently question-begging. For suppose that the mind is identical to something material. If that were true, then when Descartes claims he can doubt the existence of anything material, then since the extension of 'the mind' (per hypothesis) is something material, it necessarily follows that Descartes can doubt the existence of his mind. Hence, when construed extensionally, the very claim that Descartes can doubt the existence of anything material but cannot doubt his mind's existence presupposes from the get-go that his mind is immaterial, which is the very thing he aims to demonstrate. Therefore, when construed extensionally, Descartes' premise (1) is question-begging.

But if the intentional objects of Descartes' doubt are construed intensionally, then arguably the argument is invalid, since the clusters of

descriptions within subjects' minds are not real, intrinsic properties of the extramental objects themselves. This is arguably why we cannot infer the non-identity of the morning star and the evening star merely from the fact that, intensionally, one can know the morning star exists without knowing the evening star exists.

Therefore, Descartes' argument is either question-begging or invalid. Either way, the argument is unsuccessful.

Criticism Two: The dubitability of the mind's existence

Descartes holds that he cannot doubt the existence of his mind. But this is implausible.

Descartes' motivation for this is that doubting is a kind of thinking, and since thought presupposes a thinker to which the thought belongs, Descartes must be a thinking (and existing) subject. But it is not immediately evident that, necessarily, every thought must belong to a unified, thinking subject. It is at least conceivable that there exists a 'Platonic Realm' wherein thoughts just 'float,' as it were, not possessed by any unified, thinking, conscious subject.[70] Moreover, there seems to be nothing inherently contradictory about this state of affairs. But if this is conceivable and at least not contradictory, then surely it is not *certainly and indubitably true* that, necessarily, thoughts must belong to a unified, thinking subject. But if that is true, then Descartes' claim that his mind is indubitably existent is false.

2.2.3 Divisibility: Evaluation

Criticism One: Question-begging

Suppose that the mind is, in fact, identical to the brain. If that's the case, then it is simply false that the mind is indivisible – precisely because the brain is divisible. But this means that the very affirmation that the mind is indivisible amounts to the claim that the mind is not identical to the brain – and that is precisely the *very issue in question*. So, Descartes presupposes that the mind is not identical to the material brain (and/or body) in his very

affirmation that the mind is indivisible. Hence, this very affirmation amounts to question-begging against the physicalist.

Descartes may respond that, although technically such a claim amounts to a denial of physicalism, the claim is validated by simple phenomenological inspection. For, according to Descartes, he apprehends the mind to be metaphysically simple.

But all Descartes has shown is that he is *not aware of any parts composing his mind*. But this is crucially different from and does not entail that he is *aware that no parts compose his mind*. Merely from the fact that, say, I am not aware of any neutrinos passing through me right now, that is crucially different from and does not entail that I am aware that no neutrinos are passing through me right now. But all Descartes has shown is the former, whereas his rejoinder to the question-begging allegation needs to establish the latter.[71]

Criticism Two: Is the mind indivisible?

It is at least unclear whether the mind is indivisible. We know that severing the corpus callosum (the thick bundle of fibers that connects the two brain hemispheres) of epilepsy patients can, under certain experimental conditions, result in seemingly disunified, disparate perceptual experiences, desires, and even beliefs. It is unclear, therefore, that the mind is, after all, utterly indivisible.

The unity of consciousness as a difficulty for materialism

Despite the above criticisms, Descartes is nevertheless pinpointing a potential difficulty for materialist accounts of the mind. Consider your present experience. You see a paper in front of you, feel its texture on your hands, and hear the rustling noise it makes. Neuroscience, though, has revealed that discrete neurophysiological events in the brain process each aspect of your experience (the shapes, sizes, colors, textures, sounds, motions, depths, and so on). "Yet," writes philosopher Edward Feser:

The experience you are having is neither an incoherent jumble of distinct and disconnected features (pages, ink, motion, colors, etc.) nor is it a collection of distinct and disconnected experiences of distinct and disconnected features; it is a single, unified experience… The experience has a coherent significance or meaning, and significance or meaning for a single subject of experience. You are not only aware of the shape, texture, colors, etc. as separate elements, but are aware of them as a book; and it is you who are aware of them, rather than myriad neural events somehow each being "aware" of one particular aspect of the book (Feser (2007).

Despite the fact that the various aspects of our thoughts, perceptions, and so on are separately encoded by distinct neurophysiological processes, we nevertheless possess an integrated, unified, holistic experience rather than an awareness of a meaningless sequence of shapes, colors, motions, sounds, and textures. But how is this possible if the discrete neurophysiological processes exhaust all the facts of experience and thought? From whence comes this binding, this meaningful, continuous, unified conscious experience?

It clearly won't do to posit a higher order neural mechanisms that integrates the information from lower order neural mechanisms solve the problem. "For now", Feser points out, "all the relevant information would have to be gathered together in *this* mechanism, which itself would be composed of yet further distinct neural processes encoding distinct aspects of the visual field, and the binding problem would arise again at a higher level" (Feser (2007)).

This is not presented as an insuperable, knock-down problem, but it nevertheless seems to carry *prima facie* weight as an explanatory difficulty for materialist accounts of consciousness.

2.3 The Interaction Problem

The interaction problem for dualism consists in two seemingly intractable questions concerning the causal influence between mind and body:

(1)   *How* does the mind (body) exert causal influence over the body (mind)?

(2)   *Where* does the mind (body) causally interact with the body (mind)?

The problem seems especially intractable when we consider the radical differences between the two substances posited by Descartes' account. The mind, according to Descartes, is utterly unextended, non-spatial, and lacks energy, momentum, mass, size, and dimensions. How on earth could something lacking any spatial dimensions, lacking any energy, and being wholly non-physical causally interact with and act upon something that is essentially spatially extended, energetically influenced, and wholly physical? How could there be any causal contact between these two vastly different realms? The connection seems to be wholly inexplicable.

A difficulty also arises in considering *where* this supposed interaction is occurring. Is it occurring in the brain? If so, where? And why *there* as opposed to two picometers to the left or right? And in virtue of what does this unextended substance 'latch onto' something that is purely extended? How could it even be meaningful to talk about a dimensionless, unextended thing having a spatial location of its interaction?

These explanatory difficulties, one may argue, provide us with good reason to reject Descartes' substance dualism.

2.3.1 A Pathway to remove some mystery

To see why the interaction problem is not as mysterious as we might initially think, consider physical causation. What *is* physical causation? I aver that there is a deep mystery here, a mystery that parallels dualism's interaction problem.

Just think about it. What is this mysterious relation of 'production' whereby one object (event) causes another object (event) to exist (happen)? How does that even occur? We can specify the shapes of the objects, their spatial contiguity, their color, their relative velocities, their kinetic and potential

energy, and so on, but where is causality in this specification? In which of these facts does causation consist? To which of these facts could we point?

We say that one billiard ball causes another billiard ball to move. But what is this relation of causation? It won't do to analyze it in terms of 'production', 'power', 'capacity', or 'generation,' since these terms *are themselves causal notions* and thereby presuppose the very thing to be analyzed. But if we make no reference to such causal notions, instead referring to wholly acausal facts, then it seems (one may argue) we haven't *analyzed* causation but rather have *eliminated* it.

For instance, we may analyze causation as constant conjunction, or spatial contiguity, or some other account. But arguably, these accounts merely *eliminate* causation rather than illuminate its nature. We could analyze physical causation in terms of momentum transfer. But what is momentum transfer? Momentum is a quantity posited to account for conserved quantitative aspects of the *causal effects* of physical bodies on other physical bodies. If anything, momentum transference either *presupposes* a causal production of an effect (in which case it is not an illuminating analysis of causation), or it amounts merely to a constant conjunction in which a given quantity is conserved (in which case it seems to *eliminate* rather than *illuminate* the causal relation – which is one of the central arguments against Humean accounts of causation).

But if physical to physical causation is just as mysterious as non-physical to physical causation, it is unclear what the problem for dualism is. Indeed, many philosophers hold (due to the considerations adduced above) that causation is *primitive* – a basic, irreducible, unanalyzable metaphysical feature of reality. Indeed, these difficulties in reducing (ontologically or conceptually) the causal relation are precisely what we would expect if the causal relation is primitive.

In fact, there seems to be consensus that causal reductionist theories do not succeed. As two foremost experts on causation point out:

After surveying the literature in some depth, we conclude that, as yet, there is no reasonably successful reduction of the causal relation. And correspondingly, there is no reasonably successful conceptual analysis of a philosophical causal concept. No extant approach seems able to incorporate all of our desiderata for the causal relation, nor to capture the wide range of our causal judgments and applications of our causal concept… The prospects of a relatively simple, elegant and intuitively attractive, unified theory of causation, whether ontological reduction or conceptual analysis, are dim (Paul & Hall (2013)).

So, it seems we have *some* reason to be primitivists about causation. But if that is true, then physical-physical causation seems to be every bit as 'mysterious' as non-physical-physical causation. Under such an account, both are equally primitive, irreducible, and basic. At the very least, defenders of the interaction problem as an objection to dualism owe us an account of causation that would break the symmetry here.

A final note in relation to the interaction problem concerns the very nature of *how* questions. Usually, when we pose a *how* question, we seek a *mechanistic* explanation of how one phenomenon gives rise to or interacts with another phenomenon. But, by definition, the immaterial mind is not some *mechanistic* thing with parts arranged and operating in mechanical ways. But if that is true, then demanding a *how* explanation in relation to the mind's activity is (one may argue) a *category error*.

3 Armstrong's identity theory

In this section, I outline Armstrong's central argument for his identity theory of the mind. I then critically evaluate the plausibility of this theory.

3.1 Armstrong's argument

Armstrong's theory (presented in his article "The Nature of Mind") is a version of reductive materialism according to which (i) only physical, material things (objects, properties, relations, etc) exist, and (ii) mental

properties, states, and events (in particular) are identical to physical processes, states, and events of the physical brain (in particular).

The primary motivation Armstrong adduces in support of this view is that the progress and tremendous success of science gives us reason to think that a complete account of humans can and will be given in purely physico-chemical terms. The tremendous successes of molecular biology and neurophysiology, in particular, along with their progress and trajectory, make it very likely that the complete account of the human person will be wholly physical and chemical.

3.2 Critical appraisal

Criticism One: Neuroscience does not tell in favor of Armstrong's hypothesis

Rather than validating Armstrong's hypothesis that neuroscience will give a complete account of the human person in purely physico-chemical terms, arguably neuroscience has reinforced the inadequacy of such an account *in principle*. For, despite the success and progress of neuroscience, we are not closer to solving the hard problem of consciousness: why are *any* of those neural firings, ion movements, chemical transmissions, and so on associated with *any* qualitative, subjective, first-personal states?

Indeed, the more we study neuroscience, the more intractable the problem seems. We can specify as many of the facts as we want about the velocities, positions, momentums, sizes, shapes, colors, masses, charges, textures, movements, and so on about the neurons, molecules, and atoms in the brain, but it seems there will always be the further question: why is *any of that* accompanied by inner, subjective, qualitative experience?

Criticism Two: Question-begging

Suppose the mind is, after all, immaterial. If that is true, then science is methodologically inept to determine its intrinsic nature and existence. This is because science is inherently restricted to the investigation of publically accessible, third-person, physical things. But that means that anything

immaterial, private, and first-person is beyond the scope of scientific methodology. But if that is true, then science could not possibly arbitrate on the nature and existence of the mind. It follows, then, that the very claim that science can and will arbitrate on the nature and existence of the mind presupposes that the mind is not immaterial (i.e. is material). But that is the very thing Armstrong aims to establish by means of his argument. Hence, he seems to beg the very question at issue.

Criticism Three: Rasmussen's counting argument contra identity theory

Philosopher Joshua Rasmussen provides a forceful counting argument against materialism. Premises one and two are from Rasmussen (2015), while the other premises are my additions.

1. For any class of physical properties, the *p*s, there could be a mental property of *thinking that the ps are physical*.
2. There are more classes of physical properties than there are physical properties.
3. So, there could be more mental properties than physical properties.
4. If the nature of mental properties were identical to the nature of physical properties, then there could *not* be more mental properties than physical properties.
5. So, the nature of mental properties is not identical to the nature of physical properties.
6. If (5) is true, then Armstrong's identity theory is false.
7. Therefore, Armstrong's identity theory is false.

To motivate premise one, it is useful to consider an example. Consider *shapes*. "Given any class of shapes," writes Rasmussen, "we can coherently describe a unique mental property in terms of those shapes. There is a procedure for doing so: for any class of shapes, the *q*s, there is a distinct mental property of *thinking that the qs are shapes*." Similar procedures can be run for other physical properties and, importantly, physical properties *as such*. In this way, we actually have a

coherent *procedure* for identifying a unique mental property for any class of physical properties, so this seems not to be a matter of merely begging the question against the physicalist by a premise without independent procedural support.

Premise (2) is simply Cantor's theorem: for any particular items, the classes of those items have a greater cardinality than (i.e. outnumber) the items themselves. From this, it follows that the classes of physical properties outnumber the physical properties themselves. Premises (4) and (6) seem reasonably self-evident.

We therefore seem to have reason to reject Armstrong's identity theory on the basis of the counting argument. What's more, it seems that the counting argument lends support to some form of dualism, since it shows that the nature of mental properties cannot be identical to the nature of anything physical – precisely what we would expect under dualism but *prima facie* surprising given physicalism.

4 Ryle's behaviorism

In this section, I outline Ryle's central argument for logical behaviorism. I then critically evaluate the plausibility of this theory.

Before going further, however, a few clarifications are in order. First, behaviorism makes use of the distinction between dispositional and categorical properties. So, for the behaviorist, 'Elijah wants an apple' means something like: if you ask Elijah 'do you want an apple?', he will respond 'yes'; if Elijah has apples in the kitchen, he will go into the kitchen and get one; if you offer Elijah either an apple or a peach, he will choose the apple; and so on. So, to be in the mental state of *wanting an apple* is *either* to exhibit apple-wanting behavior *or* to have a disposition to exhibit apple-wanting behavior.

Second, there are a number of motivations one may cite in its defense (in addition to Ryle's central argument). For starters, it resolves in one fell swoop a number of problems within the philosophy of mind. The problem of other minds[72] becomes much less mysterious because there is no

difficulty in accounting for our knowledge of other people's *behavior*. The mind-body problem is much less mysterious because the interaction between the physical body and its behavioral manifestations is very well understood. The problem of mental causation, too, is much less mysterious since there is (allegedly) no difficulty in understanding how physical behavior causes other kinds of physical behavior.

4.1 Ryle's argument

Ryle's argument, which derives from his book "The Concept of Mind," is that Cartesianism with respect to the mind is a category error and that mental states are nothing but behavior or behavioral dispositions. His argument can be formulated (roughly) as follows:

1. Talk about mental concepts is correctly analyzed in terms of behavior and dispositions to behavior.
2. If (1) is true, then we should not postulate the existence of distinctively mental states, events, etc.
3. So, we should not postulate the existence of distinctively mental states, events, etc.

As noted previously, two advantages of this theory about the nature of the mind are (i) it solves the problem of other minds (since there is no mystery in coming to know the publically accessible behavior/behavioral dispositions of others); and (ii) it is arguably simpler than rival theories insofar as refrains from postulating another type of thing in its ontology (namely, private, first-person, subjective mental states).

4.2 Critical appraisal

Moorean shift

The principal objection to behaviorism is that it seems to straightforwardly contradict an inward inspection of our own minds. We *just know* that minds are more than mere behavior – we can *see* this with direct and unfaltering certitude. We can (and often do) peer into the inner,

qualitative character of our mental states. In doing so, however, we thereby falsify behaviorism. But surely, the behaviorist may note, this is just dogmatism – isn't this inner, private, what-it-is-like character of mental states the very thing the behaviorist is calling into question?

On the contrary, it is far from dogmatic, as we may employ one as-yet-unexplored tool: the Moorean shift. Any argument for behaviorism only shows that some set of premises (on the one hand) and the claim that you have private, internal mental states (on the other hand) are jointly inconsistent. But this is just to say that *either* one or more of the premises is false, *or* the claim that you have private mental states is false. But now we can ask ourselves: which is more plausible, that *at least one of the premises is false* or that *you have private mental states*? Just as for G.E. Moore there could be no question that the claim 'I have hands' is more plausible than the premises in a complex and convoluted argument for external world skepticism, there also can be no question that the claim 'you have private mental states' is more plausible than the premises in any argument for behaviorism.

Intentionality

A second problem for behaviorism is accounting for intentionality. Intentionality is the *aboutness* or *directedness toward some object or end* that is characteristic of many mental states. But how can mere behavior or behavioral dispositions be *about* anything? The behaviorist may respond that walking towards an apple, picking the apple up, and subsequently eating the apple is a series of behaviors that constitute a desire for an apple, which in turn is *about* an apple.

But human behaviors are only *about* things to the extent that they are *results of intentions* on the part of the human agent. A human behavior is only directed towards an end provided that the end is *consciously intended* as a final outcome of the behavior. But conscious intentions are internal subjective states, which is the very thing the behaviorist sought to avoid in the first place. Thus, to preserve the intentionality of mental states, behaviorists must appeal to the very privacy they deny exists.

An example helps illustrate the crucial claim of the above argument, namely that human behavior is only directed towards an end provided that the end is *consciously intended* as a final outcome of the behavior. To see this, it is important to see that my picking up the apple has a *whole host* of behavioral and environmental effects, many of which are not the objects of my intention. So, for instance, although my behaviors ruffle leaves on the ground, my behavior is not *about* or *directed toward* the ruffling of leaves. But on the basis of behavior alone, there is no objective way to delineate between my picking up the apple as opposed to my ruffling the leaves as the intentional object of my acts.[73] Hence, in order to account for intentionality, reference must be made to conscious intentional states. Behaviorism therefore cannot account for the existence of intentionality.

Counter-examples

We saw in a previous chapter that a powerful way to evaluate claims to the effect that A and B are identical (or else co-extensional) is to see if there are cases when A is present but B is absent or when B is present but A is absent. Indeed, both of these cases apply to behaviorism. Recall that behaviorism identifies mental states with behavior or behavioral dispositions. But not only can there exist mental states without behavior or behavioral dispositions, but there can also exist behavior or behavioral dispositions without mental states. Therefore, mental states are not identical to behavior or behavioral dispositions, in which case behaviorism is false. Let us consider each of these counter-examples in turn.

First, mental states can and do have no corresponding behavior. Consider somebody who has lost the ability to manifest behavior due to serious paralysis. This individual is, unfortunately, totally paralyzed – he can move almost no parts of his body and thus cannot communicate with us or behave in any way. But surely he can and does have mental states like desires, beliefs, feelings, emotions, sensations, and so on.

The behaviorist will respond that, although this individual has mental states with no corresponding behavior, he nevertheless possesses dispositions toward behavior.

This, however, is unpersuasive. Suppose we take the glass from an earlier example and alter its molecular structure so that it can now survive great falls and impacts (like being hit with a sledgehammer) without a single scratch. In this scenario, it is not just that the glass's fragility is not being realized; rather, *the glass has lost its fragility altogether*. For an entity E to possess a disposition toward some behavior B, E must have the *potential* to manifest B. That is to say, E must in some sense have the *causal power* to bring about B.

But in the paralysis case, it seems that the individual does not even have the *potential* to manifest B (say, pain behavior) due to the physiological and biochemical constitution of his nervous system, muscular system, and so on. He lacks the relevant causal powers. He no more possesses a disposition towards, say, pain-behavior than he possesses a disposition towards flying-behavior. Hence, he thereby lacks the relevant disposition towards behavior as well. Therefore, this behaviorist response fails. We can see, therefore, that there can exist mental states without behavior or behavioral dispositions.

Second, we can clearly behave *as if* we were in a specific mental state while not actually being in that mental state (for instance, screaming horrendously, wincing, crying, etc. as though we were in excruciating pain), in which case there can be behavior without corresponding mental states. But if that is true, then behavior is not identical to mental states. Instances of sleepwalking, moreover, often involve quite complex behaviors (sometimes even cooking an egg!) without corresponding conscious awareness, intentions, desires, and so on.

Circularity

Another central problem for behaviorism is that we simply cannot paraphrase away mental states by means of a purely behavioral analysis. This is because any such attempt to analyze mental states solely in terms of behavior *will itself invoke a whole host of mental states*. To demonstrate this, consider again the case of Elijah and his desire for an apple. The behavioral analysis of 'Elijah wants an apple' goes something like: if you

ask Elijah 'do you want an apple?', he will respond 'yes'; if Elijah has apples in the kitchen, he will go into the kitchen and get one; if you offer Elijah either an apple or a peach, he will choose the apple; and so on.

But a problem now arises. For Elijah will only answer 'yes' if he *understands* the question being asked, *wants* to speak truthfully, *believes* that you are listening to him, and so on. He will go into the kitchen and get an apple only if he *believes* there are apples in the kitchen, *knows* the layout of his house so as to facilitate walking into the kitchen, and so on. He will take the apple rather than the peach only if he correctly *perceives* that you are offering him a peach and an apple, and so on.

We did not, therefore, actually provide a *purely behavioral analysis* of Elijah's wanting an apple. Our apparently behavioral analysis clearly presupposes a host of other mental states – and characterizing *those* mental states in further behavioral terms will *itself* presuppose still further mental states. The analysis is not behavioral at all.

The problem results from the fact that mental states do not act in isolation; behavior only arises when there exist a whole host of mental states in conjunction. There is thus no way to give a non-circular analysis or paraphrase of a mental state in terms of behavior.

Note, further, that this criticism is precisely one of the methods of analyzing and evaluating arguments explored in a previous chapter, since it establishes that an explanation of X (namely, an explanation or analysis of mental states solely in terms of behavior) *presupposes* the very thing in need of explanation, viz. X itself.

Circularity again

A yet further difficulty involved in any behaviorist analysis is that it is circular in another respect, since "accurately characterizing a piece of behavior itself requires reference to the mental state of which the behavior is an expression" (Feser (2019)). This is because any given behavior or cluster of behaviors can be associated with a vast number of different mental states (for instance, punching the air violently can be an expression

of sheer jubilance just as much as it can be an expression of frustration and anger). But if any given behavior or cluster of behaviors can be associated with a vast number of different mental states, then to accurately characterize a given behavior one way rather than another as constituting, say, anger instead of jubilance, one must make reference to the mental state (i.e. anger) of which the behavior is an expression. But if that is the case, then one cannot adequately analyze mental state M solely in terms of behaviors B, since to accurately characterize and interpret B *in the first place* requires reference to M.

The inadequacy of the behaviorist account

But just as a vast number of *mental states* may accompany any given behavior or cluster of behaviors, a vast number of *behaviors* can, in turn, be associated with a single mental state.

Consider the agential, intentional act of *searching for water*. There are all sorts of things an agent might do if it is searching for something. Agent A_1 may be frantically running around his kitchen, screaming, and flinging open cabinet doors in his search for water (perhaps because he just burnt his hand and wants cold water to soothe it), while agent A_2 may be calmly and sedately wandering around a forest (perhaps because he or she is camping, ran out of water, and is searching for a clear river or creak nearby).

Both agents are in these cases exhibiting the intentional mental state of *searching for water*, however their behavioral manifestations are vastly different. There is thus a dilemma for the behaviorist: either there is some feature of A_1 and A_2's behavior (i.e. bodily movements) that unites them, or they are just an arbitrary pair. Clearly, though, they are not just an arbitrary pair precisely because they are both manifestations of the same intentional mental state. But there seems to be no non-arbitrary way to unite A_1 and A_2's behavioral manifestations apart from invoking intentional mental states, especially when factoring in the fact that a further agent, A_3, may be acting extremely similar to A_1 when A_3 undergoes a fit of anger (rather than a search for water) within A_3's kitchen.

The above argument can be summarized in general terms as follows: if any given mental state can be associated with a vast number of different behaviors or clusters of behaviors, then since (per behaviorism) mental states just are behavior, it follows that there must be some behavior or aspect of behavior in common among the vast number of different behaviors or clusters of behaviors. But by the very nature of this case, since wholly different behaviors or clusters of behaviors can be associated with the same mental state, there seems to be no specific behavior or aspect of behavior in common among the vast number of different behaviors or clusters of behaviors. The only thing which unites them, it seems, is the private intention associated with the mental state – and this is precisely something to which a behaviorist cannot appeal. Behaviorism thus cannot account for the fact that a single (type of) mental state can be associated with a vast number of different behaviors.

Finally, consider that a behavior that produces sound waves only counts as *saying something* if there are particular intentions behind it (otherwise, it would be a merely accidental and meaningless string of noises). Particular words or sounds only count as meaningful insofar as *minds consider them as such*, i.e. only relative to the intentions of one or more minds. But if that is the case, then even a behaviorist analysis of *saying something* presupposes non-behavioral mental states like intentions.

Making a mess of mental predicates

Notice, further, that the behaviorist analysis of 'Elijah wants an apple' may significantly differ from the behaviorist analysis of 'Jones wants an apple.' Elijah is, we can suppose, a normal apple-wanter. That is to say, his behavior is exactly what we would expect of someone who desires an apple. But suppose that Jones is on a special diet that explicitly prohibits apples. He may really, really want an apple (indeed he may love apples). However, he knows he has little willpower and thus actively avoids apples at all costs. If you ask him 'do you want an apple?' or even 'do you like apples?', he will respond with an emphatic 'No!' in order to keep apples completely out of his mind. If you hand him an apple and a peach, he will hit the apple out of your hand so as to stay true to his diet.

In this case, however, Jones does everything we would expect from someone who despises apples. His behavior in relation to wanting an apple is the exact opposite of Elijah's. But if mental states are just behavior (per behaviorism), then since the behavioral analysis of 'Elijah wants an apple' and 'Jones wants an apple' are so different, it seems that the meaning of '___ wants an apple' can be given only with reference to particular individuals (Baker (2014)).

Predicating '___ wants an apple' of Elijah will therefore have a completely different meaning to predicating '___ wants an apple' of Jones, and so on for every particular individual. But in that case, how is '___ wants an apple' even a meaningful predicate given that it has a completely different meaning when applied to different subjects? But it seems it *is* a meaningful predicate. Behaviorism thus seems to make a total mess of mental predicates.

Determinacy of mental states and indeterminacy of behavior

Another criticism of behaviorism is that "behavior is indeterminate with respect to the conceptual content one might attribute to it," whereas some mental states have a determinate conceptual content (Feser (2019)).

Take, for instance, Quine's famous example, 'gavagai'. Suppose you are a linguist attempting to translate the language of a group of people. One of the members of the group points toward a rabbit and states, 'gavagai!'. You are naturally inclined to think 'gavagai' translates to something like, 'Lo, a rabbit!'.

But mere facts about behavior allow for incompatible assignments of conceptual content to 'gavagai'. For instance, the mere fact of the language user pointing to a rabbit and stating 'gavagai' is insufficient to determine whether the language user means 'rabbit', or 'undetached rabbit part', or 'temporal stage of a rabbit', or 'potential dinner', or 'fluffy thing I saw yesterday'.

Each of these conceptual contents attributable to 'gavagai' is perfectly consistent with the facts about the language user's behavior. There is

nothing, therefore, in the behavioral facts *alone* that determines exact or unambiguous conceptual content among a range of different, incompatible conceptual contents. Hence, behavior is indeterminate with regard to its conceptual content.

However, some mental states have determinate conceptual content (for instance, consider our concept of 'triangle' which does not admit of a range of different, incompatible assignments of conceptual content). Therefore, some mental states cannot just be behavior, and hence behaviorism is false.

Epistemological argument

A final argument against behaviorism is epistemological in nature:

1. If mental states are identical to behavior or behavioral dispositions, then the only way to come to know one's own mental state is through coming to know one's own behavior or behavioral dispositions.
2. If the only way to come to know one's own mental state is through coming to know one's behavior or behavioral dispositions, then one cannot know one's own mental state without consulting one's behavior or behavioral dispositions.
3. But one can know one's own mental state without consulting one's behavior or behavioral dispositions.
4. Therefore, mental states are not identical to behavior or behavioral dispositions.

Consider a scenario in which you touch a hot stove for the first time in your life and, as a result, feel pain. You wince, cry, recoil, say 'ouch!', and so on. However, you do not first need to consult your behavioral manifestations and, as a result, come to the conclusion that you are in pain. Rather, you know that you are in pain *independently* of your behavioral manifestations. It is not by first recognizing that you are crying, wincing, recoiling, and so on that you come to realize that you are in pain; rather, you know you are in pain without consulting this cluster of behaviors.

Suppose instead that, desiring to appear stoic, you restrain all these behavioral manifestations. In that case, according to behaviorism your pain would consist solely in your *dispositions* toward pain behavior. The problem, though, is that yet again you can know you are in pain without first consulting your various behavioral dispositions. Indeed, you may not even know *how* to consult your various behavioral dispositions (in fact, arguably we cannot even *observe* dispositions – or if we can, we can observe them only when they are being realized. If we see a glass smash upon being dropped, *perhaps* we observe its fragility. But if we see the glass sitting on the table, undisturbed, we can only *infer* its fragility from what we know about glasses).

In these cases, you are not exhibiting pain behavior, nor are you observing any disposition towards pain behavior, nor do you have anything from which you could infer that you have a disposition towards pain behavior (since this is the first time in your life you have touched a hot stove and hence have no prior basis upon which to infer your behavioral dispositions in relation to hot stoves). We can see from these cases, therefore, that you can know you are in pain without consulting your behavior or behavioral dispositions. Hence, premise three is on good footing.

But isn't that the masked man fallacy?

One may object that this argument commits the masked man (i.e. epistemological-metaphysical) fallacy in that it argues in the following form:

1.   If X and Y are identical, then I cannot know X without knowing Y.
2.   I can know X without knowing Y.
3.   Therefore, X and Y are not identical.

But note that, as explained in a previous chapter, this reasoning is only fallacious in *intensional contexts* when the expression representing the intentional object of the subject's belief, knowledge, desire, etc. cannot be

swapped for co-referring expressions without potentially changing the truth value of the proposition in question.

This is created, in turn, when the different co-referring expressions are present to the subject under different clusters of descriptions. In other words, the subject fails to recognize that such expressions are, indeed, co-referring because he or she attributes different clusters of descriptions to each of them (they are, as it were, presented to the subject under different *aspects* or *senses*).

For instance, when we say that 'Lois believes Clark Kent is standing next to her', Lois is the subject, Clark Kent is the intentional object (which is expressed by the written or spoken words 'Clark Kent'), and the intensional context results from the fact that the proposition contains a propositional attitude (in this case, belief) toward an intentional object which can be represented by different expressions. We cannot swap the co-referring expression 'Superman' with 'Clark Kent' in this proposition without potentially changing its truth value precisely because Lois associates different clusters of descriptions with 'Clark Kent' and 'Superman' respectively.

But if Lois Lane *did not* associate different clusters of descriptions with concepts of 'Clark Kent' and 'Superman' (i.e. if she associated them both with *identical* clusters of descriptions), then we *could* swap these co-referring expressions without potentially changing the truth value of the proposition. The fallacy, therefore, is only present when the co-referring expressions are available to the subject under *different* aspects or clusters of descriptions.

But now we can see why the epistemological argument against behaviorism does not commit the masked man fallacy. For if behaviorism is true, then the only facts that exist are public, third-person facts. But if the only facts that exist are public, third-person facts, then the only facts available to subjects to form descriptions are public, third-person facts. But if the only facts available to subjects to form descriptions are public, third-person facts, then subjects cannot form descriptions of private, first-person

facts. This is because descriptions are built out of concepts, and our concepts can never go beyond what we have encountered in sensory experience except by means of negation, combination, or abstraction.[74] Allow me to illustrate this with a few examples.

It is undoubtedly true that a significant number of our concepts result from what we have encountered in sensory experience. For instance, I have the concepts of my aunt, my dog, my house, fire hydrants, trees, cars, the sun, and so on. All of these derive from sensory experience.

However, some of our concepts seem to go beyond our sensory experiences. Take, for instance, the following concepts: immateriality, unicorns, and triangularity. Presumably, we have not experienced any immaterial things in our ordinary sensory perception. How, then, do we have this concept?

The answer seems somewhat simple: negation. We can use our intellect to form the concept of immateriality from considering our concept of materiality, itself derived from our sensory experiences of material things, and then *negating* materiality and that which is contained within it to from the concept of immateriality. So, although the concept of immateriality does extend beyond our sensory experience in the sense that we have not experienced anything immaterial, it is nevertheless *grounded* in experience in the sense that it is a negation of our concept of materiality (which, in turn, derives from our experience of material things).

Now take our concept of unicorns. We have never experienced unicorns, but yet we have a concept of them. How? Again, the answer seems rather simple: combination. We only have a concept of unicorns because our intellect can *combine* various concepts of things which we *have* experienced (like horns, horses, and so on) into a complex concept of something we have *not* experienced (like unicorns). Again, although the concept of unicorns does extend beyond our sensory experience in the sense that we have not experienced any unicorns, it is nevertheless *grounded* in experience in the sense that it is a combination of our concepts of horns, horses, and so on (which, in turn, derive from our experience).

Finally, take the concept of triangularity. No one has experienced triangularity *as such*; all we have ever experienced are particular triangles that possess individuating features that triangularity in the abstract lacks. Triangularity is, in essence, that which is common or universal to all triangles.

Despite our never having experienced triangularity, we nevertheless can conceptualize it by means of *abstraction*. We can use our intellect to form the concept of triangularity from considering particular triangles (derived from our sensory experiences of them in geometry textbooks, art, architecture, and so on) and then *abstracting* from the various particularizing and individuating features in order to grasp what is common to all triangles. Again, although the concept of triangularity does extend beyond our sensory experience in the sense that we have not experienced triangularity as such, it is nevertheless *grounded* in experience in the sense that it is an abstraction from our experiences of particular triangles.

How, though, is all of this relevant to the masked man fallacy? Well, as we saw in a previous paragraph, descriptions are built out of concepts. Our concepts, in turn, can never go beyond what we have encountered in sensory experience except by means of negation, combination, or abstraction.

Crucially, though, we have a concept of private, first-person mental states. Now, since an experience of x can only occur if x exists, and since behaviorism holds that private, first-person mental states do not exist, it follows that under behaviorism we cannot experience private, first-person mental states. Hence, our concept of subjective, first-person mental states must go beyond what we have encountered in experience. Since concepts can only extend beyond what we have encountered in sensory experience by means of negation, combination, or abstraction, it follows that under behaviorism our concept of private mental states must be either a negation, combination, or abstraction from things that we have experienced (which, for behaviorism, are *exclusively* objective, third-person, public facts).

But our concept of private mental states is not a *negation* of a concept of something objective, third-person, and public. This is because concepts that are mere negations of other concepts (the example used earlier was immateriality) seem to have conceptual content that is wholly parasitic on (i.e. wholly dependent upon and built from) the conceptual content of that which it negates. So, for instance, immateriality has conceptual content or meaning that is parasitic on the meanings attached to materiality. When we say that immaterial things are unextended, for example, this is clearly parasitic on our understanding of material things as spatially extended.

But the same does not apply to our concept of private, first-person mental states, since its meaning contains things that are not merely negative characterizations parasitic on sensible, public things but rather contains a whole host of positive, robust, non-parasitic content. For instance, one such content is the idea of 'what-it-is-like-ness', the raw qualitative 'feel' that is associated with private mental states such as the taste of chocolate or the smell of lilacs. It seems clear that this content is not merely dependent upon, built out of, or parasitic upon the negation of 'non-what-it-like-ness'. In other words, we do not understand 'what-it-is-like-ness' to be a mere negation of 'non-what-it-is-like-ness' in the way that we understand 'unextended' to be a mere negation of 'extended'. Therefore, our concept of private, subjective, first-person mental states is not merely a negation of the concept of something objective, third-person, and public.

Furthermore, our concept of private mental states is not a *combination* of concepts which are of objective, third-person, public things.

One argument for this runs as follows. Combining one concept of a public thing with the concept of another public thing can only result in a complex concept that is public. Combining two public concepts really just amounts to conceptualizing something that exists in extramental reality that is some rearrangement of already existing public things (like horns and horses in the case of unicorns, or like apes and humans in the case of bigfoot, and so on). But merely rearranging objective things seems categorically insufficient for bridging the gap between objectivity/publicity and subjectivity/privacy. Take two cubes. When you conceptualize them

taped together, say, as a single object, is that object any less public or objective than the two separate original cubes? We can see that the mere combination or rearrangement of the two cubes is insufficient for bridging the gap from purely objective, third-person, public things to subjective, first-person, private things.

Could three cubes do it? What about four? It seems that any number of publicly accessible cubes, when combined or rearranged together, could never bridge the gap to form subjectivity and privacy. Could a sphere do it? How about a rectangular prism? What if we made the cube a few inches longer in width? Could it do it then?

It seems evident that mere differences with respect to shape, color, size, spatiotemporal location, number, or type of objective thing are irrelevant here: combining and rearranging them will form only further objective, public realities. By the principle of relevant differences, then, wit seems plausible that merely combining or rearranging any number of objective things is insufficient to produce subjective things. But if combining our *concepts* of public things just amounts to conceptualizing the individual public things rearranged or combined together in reality, and if (per the above argument) the individual public things rearranged or combined together in reality are insufficient for subjectivity (meaning they are thus objective and public), then our concept of the public things rearranged or combined together in reality – if it is to be an accurate representation – must also be a concept of a public thing. Therefore, our concept of private mental states is *not* a combination of concepts which are of objective, third-person, public things.[75]

Finally, our concept of private mental states is *not* an abstraction from individual objective things. Objective, public things (like rocks, stones, trees, planets, and so on) by definition do not have subjectivity or privacy in common, since if they did they would not be objective and public things in the first place. Therefore, our concept of private mental states is *not* an abstraction from individual objective things.

From all of this, we can see that, under behaviourism, we could not possibly have obtained the concept of private, subjective mental states. As we saw above, under behaviourism our concept of private mental states must be either a negation, combination, or abstraction from things that we have experienced (which, for behaviorism, are *exclusively* objective, third-person, public facts). But as the aforementioned paragraphs indicate, our concept of private mental states cannot be a negation, combination, or abstraction from public things. Therefore, given behaviorism, we could not possibly have obtained the concept of private, subjective mental states.

But since our descriptions are confined solely to things of which we have concepts, that means our descriptions (and hence the clusters of descriptions that we associate with expressions that represent intentional objects) are solely confined to descriptions of public, third-person, objective things (per behaviorism).

Crucially, though, the *only way* that our clusters of descriptions associated with co-referring expressions could *relevantly differ* so as to create an intensional context would be if one cluster of descriptions was of private, first-person, and subjective mental states while the other was of public, third-person, objective things.

Recall that an intensional context is one in which one expression representing the intentional object is associated in the subject's mind with one cluster of descriptions, while a co-referring expression for the same intentional object is associated in the subject's mind with a *different* cluster of descriptions. But without the subjective, private, first-person cluster of descriptions that the subject associates with (say) the expression 'pain', the subject would only have public, objective, third-person descriptions with which he or she associates the mental state of pain. But the only plausible set of third-person, public descriptions the subject could associate with his or her pain is a cluster of his or her behaviors or dispositions to behavior – there are simply no other non-private, non-subjective, relevant descriptions to associate with her mental state of pain.

But if this is true, then *given behaviorism there simply is no intensional context present here*, since the only cluster of descriptions the subject associates with her mental state of pain are descriptions of behavior and behavioral dispositions. In that case, though, no intensional context exists since the cluster of descriptions the subject associates with her mental state is identical to that cluster he or she associates with her behaviors or behavioral dispositions.

So again, since behaviorism entails that we cannot have a cluster of descriptions that are subjective, private, and first-person in nature, that means the only possible way an intensional context could exist in this scenario does *not* obtain, in which case it is not an intensional context.

But if that is true, then we *can* swap co-referring expressions without potentially changing the truth value of the proposition in question, in which case the reasoning does *not* commit the epistemological-metaphysical fallacy (since the fallacy only arises in intensional contexts). If that is the case, though, then the epistemological argument succeeds in establishing the falsity of behaviorism.

Getting formal

This chain of reasoning became quite complex and abstract at times, so it is worth recapping and formally reconstructing the line of argument. The argument against behaviorism in question reasons that, since one can know his or her mental states without knowing his or her behavior or behavioral dispositions, mental states cannot be identical to behavioral dispositions.

One objection to this argument is that it commits the masked man fallacy. However, this fallacy only arises in intensional contexts. But, as we've seen, the epistemological argument does *not* involve an intensional context *given behaviorism*. Therefore, the argument does not commit the masked man fallacy.

The following argument is a formalized version of the chain of reasoning that led us to this conclusion. Brace yourself, though – it's long!

Let "objective, public, third-person" = O
Let "subjective, private, first-person" = S

Assumption: Every fact is either an O-fact or S-fact, and no fact is both an O-fact and S-fact.

1. If the masked man fallacy is present in the epistemological argument, then there is an intensional context present with one intentional object that has at least two co-referring expressions, E_1 and E_2.

2. If there is an intensional context present with one intentional object that has at least two co-referring expressions, E_1 and E_2, then:

    (1) a subject A has a propositional attitude P towards E_1;
    (2) there exists a co-referring expression, E_2 (co-referring with E_1);
    (3) A associates a cluster of descriptions, C_1, with E_1;
    (4) A associates a cluster of descriptions, C_2, with E_2; and
    (5) C_1 and C_2 are different clusters of descriptions.

3. Descriptions are built (constructed) solely out of our concepts.
4. Our concepts must derive from:

    (a) sensory experience; or
    (b) negation; or
    (c) combination; or
    (d) abstraction

5. If our concepts must derive from (a), (b), (c), or (d), then if a purported concept C derives from none of (a), (b), (c), or (d), then we do not, in fact, have a concept of C.

6.   Therefore, if a purported concept C derives from none of (a), (b), (c), or (d), then we do not, in fact, have a concept of C. (4,5)
7.   A sensory experience of x can occur only if x exists.
8.   S-facts do not exist (per behaviorism).
9.   Therefore, a sensory experience of S-facts cannot occur. (7,8)
10.   If a sensory experience of S-facts cannot occur, then our concept of S-facts cannot derive from (a).
11.   Therefore, our concept of S-facts must derive from either (b), (c), or (d). (4,9,10)
12.   If our concept of S-facts derives from (b), then the conceptual content of our concept of S-facts is wholly parasitic on the conceptual content of that which it negates (viz. our concept of O-facts).
13.   But the conceptual content of our concept of S-facts is NOT wholly parasitic on the conceptual content of that which it negates (namely, our concept of O-facts).
14.   Therefore, our concept of S-facts does not derive from (b). (12,13)
15.   If our concept of S-facts derives from (c), then the conceptual content of our concept of S-facts is just a combination of concepts which are concepts of O-facts.
16.   But the conceptual content of our concept of S-facts is NOT just a combination of concepts which are concepts of O-facts. (19,23)

Sub-argument #1

17.   If the conceptual content of our concept of S-facts is nothing more than a combination of concepts which are concepts of O-facts, then (our concept of) S-facts is wholly analyzable in terms of (our concept of) O-facts.
18.   But (our concept of) S-facts is NOT wholly analyzable in terms of (our concept of) O-facts.
19.   Therefore, the conceptual content of our concept of S-facts is NOT just a combination of concepts which are concepts of O-facts. (17,18)

Sub-argument #2

20. Combining any number of any kind of O-type objects cannot by itself result in a S-type object (by the principle of relevant differences).

21. If combining any number of any kind of O-type objects cannot by itself result in a S-type object, then combining the concepts of any number of any kind of O-type objects cannot by itself result in a concept of a S-type object (provided that our concepts adequately match reality).

22. If combining the concepts of any number of any kind of O-type objects cannot by itself result in a concept of a S-type object, then the conceptual content of our concept of S-facts is NOT just a combination of concepts which are concepts of O-facts.

23. Therefore, the conceptual content of our concept of S-facts is NOT just a combination of concepts which are concepts of O-facts. (20,21,22)

24. Therefore, our concept of S-facts does not derive from (c). (15,16)

25. If our concept of S-facts derives from (d), then our concept of S-facts is an abstraction from individual O-type things.

26. But an abstraction in respect R from individual things requires that all such individual things possess R.

27. But by their very nature, no O-type thing possesses S-facts.

28. If abstraction in respect R from individual things requires that all such individual things possess R, but no O-type thing possesses S-facts, then our concept of S-facts is NOT an abstraction from individual O-type things.

29. Therefore, our concept of S-facts is NOT an abstraction from individual O-type things. (26,27,28)

30. Therefore, our concept of S-facts does not derive from (d). (25,29)

31. Therefore, our concept of S-facts derives from none of (a), (b), (c), or (d). (11,14,24,30)

32. Therefore, per behaviorism we do not have a concept of S-facts. (5,31)

33. If descriptions are built (constructed) solely out of our concepts, but we do not even have a concept of S-facts, then all our descriptions are built (constructed) solely out of concepts of O-facts (per *Assumption*).

34. Therefore, all our descriptions are built (constructed) solely out of concepts of O-facts. (3,32,33)

35. But if all our descriptions are built (constructed) solely out of concepts of O-facts, then A's cluster of descriptions C_1 and A's cluster of descriptions C_2 are both built entirely from concepts of O-facts.

36. Therefore, A's cluster of descriptions C_1 and A's cluster of descriptions C_2 are both built entirely from concepts of O-facts. (34,35)

37. But the only plausible clusters of descriptions built entirely from concepts of O-facts that A could associate with E_1 and E_2 are *identical* clusters of descriptions (namely, descriptions of behavior and behavioral dispositions).

38. If A's cluster of descriptions C_1 and A's cluster of descriptions C_2 are both built entirely from concepts of O-facts, and if the only plausible cluster of descriptions built entirely from concepts of O-facts that A could associate with E_1 and E_2 are identical clusters of descriptions, then C_1 and C_2 are identical clusters of descriptions.

39. Therefore, C_1 and C_2 are identical clusters of descriptions. (36,37,38)

40. If C_1 and C_2 are identical clusters of descriptions, then (5) is false.

41. If (5) is false, then there is NOT an intensional context present with one intentional object that has at least two co-referring expressions, E_1 and E_2. (2)

42.   If there is NOT an intensional context present with one intentional object that has at least two co-referring expressions, E_1 and E_2, then the masked man fallacy is not present in the epistemological argument. (1)

43.   Therefore, the masked man fallacy is not present in the epistemological argument. (40,41,42)

Whew! Although the chain of reasoning may seem complex, it becomes much more manageable when considered in small chunks, step-by-step, and in a slow, reflective manner.

Even if this argument doesn't succeed, though, behaviorism is still stuck with a seemingly insurmountable epistemological problem in accounting for our knowledge of pain. We can be in the mental state of pain (and hence know we are in pain) without exhibiting any pain-behaviors whatsoever, as in the case when we seek to hide our pain. If pain *just is* certain behaviors or behavioral dispositions, then how could we come to know we are in pain in a scenario where we not only do not manifest any pain-behaviors, but also cannot know we have a present disposition towards pain? At the very least, the epistemological argument is extremely puzzling under behaviorism.

Behaviorism summary

First, we saw that the Moorean shift can be employed against behaviorism. Any conjunction of premises cited in favor of behaviorism will be less plausible than the simple premise that 'I have private, subjective mental states', in which case at least one of the premises in any argument for behaviorism must be false.

Second, behaviorism cannot account for intentionality since mere behavior or behavioral dispositions cannot be *about* anything.

Third, mental states can and do have no corresponding behavior, in which case mental states cannot be identical to behavior. Moreover,

behaviors can and do have no corresponding mental states, in which case behavior cannot be identical to mental states.

Fourth, it is simply impossible to paraphrase away mental states by means of a purely behavioral analysis, since any such analysis implicitly presupposes mental states. Moreover, accurately characterizing a behavior in the first place requires reference to the mental state of which it is a manifestation, in which case mental states cannot be analyzed purely in terms of behavior.

Fifth, the behaviorist account of mental states is inadequate because there is no non-arbitrary way to unite different subjects' vastly different behavioral manifestations of a single type of mental state apart from invoking intentional mental states. What's more, a behaviorist analysis cannot even be given of something as simple as *saying something*, since this presupposes non-behavioral mental states like intentions.

Sixth, behaviorism entails an absurd position in relation to mental predicates since a single mental predicate often has completely different meanings when applied to different individual.

Seventh, behavior is indeterminate with respect to conceptual content, whereas some mental states are determinate with respect to conceptual content, in which case not all mental states are identical to behavior.

Eighth, since we can know our mental state without knowing our behavior or behavioral dispositions, it follows that mental states cannot be identical to behavior or behavioral dispositions. It was charged that this argument commits the masked man fallacy. But because behaviorism entails that any cluster of descriptions a subject could plausibly associate with co-referring expressions in relation to his or her intentional object would be identical (namely, descriptions of behavior and behavioral dispositions), an intensional context is not present. And since an intensional context is not present, no such fallacy is committed.

Let's turn to our final theory concerning the nature of the mind: eliminativism.

5 Churchland's eliminativism

In the following sections, I outline Churchland's central arguments deriving from his "Matter and Consciousness."

Before moving forward, it will be useful to understand what eliminative materialism (i.e. eliminativism). Eliminativism holds that (i) the only events, properties, objects, and so on that exist are material, and (ii) *folk psychology* is a false theory that will ultimately be replaced by a mature neuroscience.

Folk psychology is the conceptual apparatus we normally use to predict, describe, and explain the behavior of human beings and other minded creatures. For instance, this is a folk psychological explanation: Jones walked towards the water fountain because he *desired* a drink. Folk psychology characteristically employs such notions as desires, beliefs, pains, feelings, wishes, thoughts, and other mental terms.

5.1 From explanatory poverty

1. Folk psychology is afflicted by widespread and long-standing failures in explanation and prediction.
2. An approach afflicted by longstanding and widespread failures in explanation and prediction should be rejected.
3. Therefore, folk psychology should be rejected.

5.2 Pessimistic induction from conceptual history

1. Most of our past folk frameworks (theories, concepts, etc.) have been wrong.
2. If most past folk frameworks have been wrong, then probably folk psychology is wrong.
3. Probably, folk psychology is wrong.

5.3 From utility

1.   There are so many more ways that neuroscientific explanations may proceed if they are not required to conform to folk psychological theories (i.e. eliminative materialism has high utility for neuroscience).
2.   If (1) is true, then eliminativism is likely true.
3.   Therefore, eliminativism is likely true.

5.4 Critical appraisal

In the following sections, I critically evaluate Churchland's three arguments. The purpose, once again, is not so much the content of the analysis and evaluation; rather, the focus is the *methods, tools, and skills* that we are developing and applying.

5.4.1 From explanatory poverty

Criticism One: The explanatory adequacy of folk psychology

Premise one seems false. Consider the following explanation:

Joe *desires* to apply the methods and tools from previous chapters, and he *believes* that critiquing Churchland's argument from explanatory poverty is one way to help ensure that this desire is fulfilled. He *thinks* that typing on the computer will facilitate this critique, and then *decides*, on the basis of his *thoughts, beliefs,* and *desires*, to type on his computer.

This seems perfectly adequate as an explanation of my behavior. Now, it may not be a *complete* explanation, but that isn't its intent in the first place. It is not a necessary condition for an explanation's success that it be a *complete* account of the *entirety* of some phenomenon.
In general, the explanatory deficiencies Churchland pinpoints (explanations of sleep, mental illness, and so on) simply lie *outside the scope* of folk psychological explanation. But a framework's inability to explain something that is inherently beyond its scope is not an explanatory deficiency on the part of the framework, any more than (say) general

relativity's inability to explain the origins of domestic terrorism counts as a deficiency in general relativity.

Criticism Two: Presupposing folk psychology is a theory

Churchland's argument presupposes that folk psychology is a theoretical framework rather than the very data for which any theoretical framework must account. But without justifying this presupposition, Chruchland's argument is itself unjustified.

More importantly, this assumption seems implausible. It seems that beliefs, desires, pains, wishes, hopes, and thoughts *are the very data that any theory of mind needs to explain*. It seems implausible that (say) we come to recognize some phenomenon in need of explanation (such as inner states), and, upon trying to explain or account for such a phenomenon, posit theoretical entities called 'beliefs,' 'desires,' 'thoughts,' and so on to explain and/or predict the raw data. It seems, rather, that such concepts simply denote the very data for which any theory must account.

5.4.2 Pessimistic induction from conceptual history

Criticism One: A relevant difference between past folk theories and folk psychology

Churchland presupposes that past folk theories are relevantly similar to folk psychology so as to legitimately facilitate an inductive generalization from their failure to folk psychology's probable failure. We therefore need a symmetry breaker if we want to avoid Churchland's argument. But there do seem to be relevant differences between past folk theories and folk psychology that delegitimize this inductive generalization.

One such relevant difference is that past folk theories almost invariably involved concocting some unobservable explanatory account or positing unobservable entities that are clearly beyond our ordinary and direct awareness of things. For instance, phlogiston is an entity that clearly was not observable. The celestial spheres of ancient cosmology were well beyond our ordinary and direct awareness. Caloric was a theoretically

postulated fluid that underlies the transmission of heat, and while the heat *itself* was directly experienced, the caloric fluid was not and was invoked as a theoretical posit.

But folk psychology seems relevantly dissimilar to these folk theories in that it concerns objects of our direct, immediate, unfaltering awareness. Folk psychological notions concern our inner, directly accessible states rather than indirect, unobservable postulations in extramental reality. But given that direct awareness and immediate access is a relevant difference between folk psychology and past folk theories, it seems that the inductive inference from the failure of past folk theories to folk psychology's failure is blocked.

Criticism Two: Inter-theoretic entity preservation

A general lesson from conceptual history is that, when a more sophisticated scientific theory replaces a false folk theory, there was nevertheless inter-theoretic entity preservation. So, although we come to have *a different understanding* of some entity or phenomenon, the *entity or phenomenon itself* is (in general) still taken to exist. We did not jettison our belief in the existence of stars upon replacing our folk conception of stars; we did not jettison our belief in the existence of space upon replacing our folk conception of space; we did not jettison the very existence of heat upon replacing our folk caloric theory; and so on.

But given such inter-theoretic entity preservation, Churchland's inductive generalization is far too hasty. He is unwarranted in inferring that, likely, our folk psychological entities don't exist; rather, he is only entitled (at best) to infer that our *understanding* of such entities will be altered. Hence, his *desired* conclusion (namely, that there likely are no such things as desires!) does not follow and is actually contravened by the conceptual history of inter-theoretic entity preservation.

Criticism Three: Self-defeat

Arguably, Churchland's argumentative structure is self-defeating. For consider the following argument:

1. Most past theories concerning the nature of the mind have been wrong.
2. If most past theories concerning the nature of the mind have been wrong, then eliminative materialism is likely wrong (since eliminative materialism is a theory concerning the nature of the mind).
3. Therefore, eliminative materialism is likely wrong.

In the absence of symmetry breakers between Churchland's pessimistic induction on past folk theories (on the one hand) and this argument (on the other), Chruchland's argument is self-defeating.

5.4.3 From utility

Criticism One: Non-sequitur

Arguably, it simply does not follow that x is likely true from the fact that x is useful to a discipline. In population genetics, it is often very useful (i) to model populations as having infinitely many members, and (ii) to assume that populations are in Hardy-Weinberg equilibrium; in oceanography and fluid dynamics, it is often very useful to assume (i) that the ocean is infinitely deep, and (ii) that fluids are continuous substances not composed of discrete, particulate sub-units; and so on. Nevertheless, every single one of these assumptions is not even approximately true. Utility, then, doesn't seem to track truth in the robust way required for Churchland's argument. Therefore, merely from the fact that eliminative materialism may be useful to neuroscientific progress, we have little to no reason to think eliminative materialism is true.

Criticism Two: The essentiality of folk psychology to neuroscience
Moroever, arguably Churchland's argument is that eliminative materialism actually seems to remove the very possibility of neuroscience. The reliability, success, and progress of neuroscience crucially relies upon the existence of and reliable access to subjective, inner, qualitative, common sense psychological states.

As Feser puts it,

> The relevant micro-level phenomena uncovered by science cannot even be identified or understood without constant reference to the commonsense macro-level phenomena they underlie, so that the latter cannot coherently be eliminated in favor of the former. Hence, consider any claim that some mental phenomenon M is correlated with some brain process B, and ought to be eliminated from our ontology altogether and replaced by B. For any such argument to get off the ground, we first have to be able to identify B, as opposed to some other neural process, as the relevant process (Feser (2019)).

But the only way to know that M is correlated with B as opposed to some other brain process is to rely upon the introspective, commonsense report of the person whose brain we are studying. If the neuroscientist is wrong about his or her presupposition of the reliably accurate introspective report of the person's own mental state, then the very evidential foundation upon which the neuroscientific, correlative evidence is based seems to collapse.

Once again, this isn't offered as a knock-down argument; it is an invitation to further thought, reflection, and discovery.[77]

In the next and final section, we will explore a potentially groundbreaking argument for the existence of other minds.

6 The problem of other minds

Before delving into the following argument, it will be helpful to clarify the relevant notion of *probability* at play. Once again, probabilities are understood as reasonable degrees of expectation (i.e. epistemic probability). In a probability analysis which employs epistemic probabilities, we are concerned with the degree to which we *expect* an outcome *given our justified acceptances*.

With the aforementioned clarifications out of the way, we can turn now to a fascinating argument (leveled by Pruss (2019) for the conclusion that your

mind is not the only mind in existence. In what follows, I will both explicate and expand upon Pruss' argument.

In the following argument, 'humans' will refer to physically (including neurologically) healthy mature humans. Now, suppose there are n humans. Let $Q_1, \ldots, Q_n$ be their non-mental qualitative profiles: complete descriptions of their non-mental life in qualitative terms. Let H_i be the hypothesis that everything with profile Q_i is conscious. For instance, if we consider human #1, Q_i is a complete description of this human's non-mental life (behavior, physiological state, third-person properties, etc.) in qualitative terms. H_1 is, then, the hypothesis that anything with human #1's non-mental profile has conscious, qualitative, subjective inner experience – a mental life.

Now consider the following three hypotheses:

> M: All humans have a mental life.
> N: Exactly one human has a mental life.
> Z: No human has a mental life.

Assume further that our background information includes the fact that there are at least two humans.

Given this background information, the hypotheses are clearly mutually exclusive. It cannot be the case that, say, all of the $\geqslant 2$ humans have mental lives (hypothesis M), while at the same time one and only one human has a mental life (hypothesis N) – for that would entail that all humans and not all humans have mental lives, which is clearly absurd. Similar reasoning reveals that, given the background information, the three hypotheses in question are mutually exclusive.

Now, add that there are n humans on earth, where n is in the billions. Add further that they have profiles $Q_1, \ldots, Q_n$, all of which are different. What's a reasonable thing to think now?

Well, N is no more likely (in terms of prior probability) than M or Z. For one thing, both Z and M are categorically uniform in that they treat *similar non-mental profiles* alike with respect to *possessing mental life* (thereby maintaining a continuity or uniformity across a category of similar things). This categorical uniformity adds to their simplicity and therefore intrinsic probability, whereas N 'builds in' a break in uniformity/continuity and increases the complexity of the hypothesis by means of an arbitrarily limited number of minded Q_i profiles (viz. 1 as opposed to 2 or 3 or 365).

For another thing, the principle of indifference (when applied to prior probabilities) dictates that we assign equal prior probabilities to hypotheses when (i) relevant evidence is absent that could serve to raise the probability of one as opposed to another, and (ii) there are no internal, hypothesis-specific reasons (e.g. coherence, simplicity) to favor one hypothesis over another. Since we are considering prior probabilities, we are (per the nature of the situation) not taking into account certain evidential considerations. So, condition (i) is met for this case. Moreover, we have seen that internal, hypothesis-specific reasons actually count *against* N and *in favor of* M and Z, as N is less simple than M and Z. Even on the supposition that such internal, hypothesis-specific reasons do *not* count against N, we must nevertheless assign a prior probability of ⅓ to N per the principle of indifference. Hence, N has a maximum prior probability of ⅓ (although, again, we have good reason to hold it is less intrinsically probable than this).

To re-cap: although we have reason to hold that N is less intrinsically probable than M and Z, we can (for the sake of maximal conservatism) suppose that all three hypotheses are equally likely. Hence, all have a maximum prior probability of ⅓. Furthermore, if N is true, exactly one H_i is true.

Moreover, all the H_i are just about equally on par given N (per the principle of indifference). Because of this, the probability that H_i is true given that N is true is 1/n. This is because there are *n* humans, and on the supposition that N is true, only one of them is minded – and since all such humans are equally on par in their prior probability of being that one lucky

minded individual (per the principle of indifference), it follows that the probability that any given human having a mind (H_i) on the supposition of N is 1/n. To formalize this a bit:

$$\text{For each i, } P(H_i|N) = 1/n$$

where $P(H_i \mid N)$ stands for *the probability of H_i given N.*

From the above, we can deduce that $P(H_i \& N)$ is at most about $1/(3n)$. To see this deduction, it is helpful to see that the following equation can serve as a definition of conditional probability:

$$P(A|E) = \frac{P(A \& E)}{P(E)}$$

Or, equivalently:

$$P(A \& E) = P(A|E) \times P(E)$$

To see why this equation successfully captures conditional probability, it is helpful to examine a concrete illustration. Suppose we have a class of 60 students. Suppose that 35 students had chicken for lunch, while 45 students had steak for lunch. Now, it clearly cannot be the case that each student only ate one type of meat for lunch, for that would entail there being 80 students (contrary to our supposition that there are 60 students). In fact, it must be the case that 20 students had both chicken and steak, 15 students only had chicken, and 25 students only had steak.

With these statistics, we can ask: what is the probability that a student ate steak? To find this, simply divide the number of students who ate steak (45) by the total number of students (60) to obtain 0.75 or 75%.

We can also ask: what is the probability that a student ate chicken *given that the student ate steak*?

Instead of dividing by the total number of students, our reference class actually becomes the total number of students who ate steak. Because of

this, we just take the number of students who *ate chicken in addition to eating steak* and divide it by our reference class (the total number of steak-eating students). For this, we obtain 20/45, or about .44 or 44%.

More carefully, what we actually performed in this calculation was the probability that a student ate both chicken and steak divided by the probability that the student was in our reference class of steak eaters. In other words, we performed the following calculation:

$$P(\text{ate chicken} \mid \text{ate steak}) = \frac{P(\text{ate chicken \& steak})}{P(\text{ate steak})}$$

Plugging in the requisite numbers, we get:

$$P(\text{ate chicken} \mid \text{ate steak}) = \frac{\dfrac{20}{60}}{\dfrac{45}{60}}$$

This becomes 20/45, as the 60's in the denominators cancel out, which in turn equals 0.44 or 44% (as before). This simple illustration should make it reasonably clear as to why the conditional probability equation is as it is.

From all of the above considerations, it can be straightforwardly deduced that $P(H_i \text{ \& } N)$ is at most about $1/(3n)$. Because $P(A \text{ \& } B) = P(A|B)P(B)$, it follows that $P(H_i \text{ \& } N) = P(H_i|N)P(N)$.

From earlier considerations, we deduced that $P(H_i|N)$ is $1/n$ and that $P(N)$ is ⅓. So, $P(H_i \text{ \& } N) = (1/n)(1/3) = 1/(3n)$. Therefore, $P(H_i \text{ \& } N)$ is $1/(3n)$.

On the other hand, however, $P(H_i|Z) = 0$ and $P(H_i|M) = 1$. This is because Z rules out any Q_i being minded, whereas M entails that all Q_i are minded.

But now suppose I learn that Q_m is *my* profile. I then learn that H_m is true. Clearly, this rules out the all-zombie hypothesis Z, and it also rules out most

of the (H_i & N) conjunctions (in fact, it rules out all of them except for one, namely (H_m & N)). From our three hypotheses on the table, then, the only two that are compatible with the new data (Q_m) are the following two mutually exclusive hypotheses:

(1) (H_i & N); and
(2) M

Crucially, though, my posterior probability (after learning Q_m) for (H_m & N) is now approximately *at most* 1/(n+1). Why is that so?

The trick is to employ the following fact: If A and B each entail E, then the ratio of P(A) to P(B) is the same as the ratio of P(A|E) to P(B|E). This is because A and (A&E) are logically equivalent when A entails E. But that means that P(A) must be the same as P(A&E). Now, recall from earlier that:

$$P(A|E) = \frac{P(A \& E)}{P(E)}$$

But since we just deduced that P(A) must be the same as P(A&E), it follows by substitution that the following is true when A entails E:

$$P(A|E) = \frac{P(A)}{P(E)}$$

By the same token, moreover, P(B|E) = P(B)/P(E), since we supposed that B likewise entails E. So, using these facts, we get:

$$\frac{P(A|E)}{P(B|E)} = \frac{\frac{P(A)}{P(E)}}{\frac{P(B)}{P(E)}}$$

And because the P(E)'s in the denominators mathematically cancel out, when A and B each entail E it follows that:

$$\frac{P(A|E)}{P(B|E)} = \frac{P(A)}{P(B)}$$

Now let's return our attention to the crucial claim, namely that my posterior probability for (H_m & N) is approximately *at most* $1/(n+1)$. Given the previous assumed background information, (H_m & N) entails Q_m and M entails Q_m. And now we have a situation identical to the situation above concerning A and B each entailing E. So, the ratio of the posterior $P(H_m\&N|Q_m)$ to the posterior $P(M|Q_m)$ is identical to the ratio of $P(H_m\&N)$ to P(M) (i.e. the ratio of their priors).

But, as we deduced earlier, $P(H_m\&N)$ is $1/(3n)$ whereas P(M) is ⅓, and the ratio can therefore be represented as follows:

$$\frac{P(H_m\&N|Q_m)}{P(M|Q_m)} = \frac{\frac{1}{3n}}{\frac{1}{3}} = \frac{1}{3n} \times 3 = \frac{1}{n}$$

Therefore, the ratio of the posterior of (H_m & N) to the posterior of M is 1:n.

But as we saw earlier, (H_m & N) and M are mutually exclusive. Their posterior probabilities, therefore, can *at most* equal 1. But when we have a ratio between x and y of the form x:y, and x and y add up to equal 1 (i.e. they are mutually exclusive and exhaustive), the following two equations must be true:

$$P(x) = \frac{x}{x+y}$$

$$P(y) = \frac{y}{x+y}$$

This may seem a bit abstract at first, so it will be helpful to concretize the situation. Suppose we have yet another class of students. Suppose further that the ratio of males to females in the class is 5:4. To simplify, suppose the class only contains 9 students in total. What is the probability that a student chosen at random is male?

Clearly, it is 5/9. Similarly, the probability that a student chosen at random is female is 4/9. In these cases, we have a ratio of 5:4 (x:y), and the probability of a random member of the class being from one group (the males) is 5/(5+4) (in other words, x/(x+y)), whereas the probability of a random member of the class being from another group (the females) is 4/(5+4) (in other words, y/(x+y)). From this concrete example, we can see with relative intuitive ease that, when there is a ratio of x:y, and when x and y cannot both be true, *at most* $P(x) = x/(x+y)$ and *at most* $P(y) = y/(x+y)$.

Hence, because the ratio of the posterior probabilities of (H_m & N) to M is 1:n, and because the two hypotheses in question cannot both be true, it follows that *at most* the posterior probabilities must equal $1/(n+1)$ for (H_m & N) and $n/(n+1)$ for M. Therefore, the posterior probability of (H_m & N) must be less than or equal to $1/(n+1)$.

And since n is in the order of billions, n is at least 1 billion, in which case the posterior probability of (H_m & N) is less than or equal to 1/1,000,000,001. Thus, the probability that my human mind is the only human mind has a probability of less than or equal to 0.000000001. Therefore, the probability that there is another human mind is greater than or equal to 0.999999999. The probability that there is at least one other mind is therefore *at least* 0.999999999.

The result is monumentally significant: we seem virtually *certain* that there exists at least one mind apart from our own mind. This, I aver, goes a substantial way towards resolving the problem of other minds.

How might we proceed from here? In particular, how might we reasonably infer that all humans have a mental life from the fact that we are virtually

certain that at least one other such human has a mind (a mind apart from our own, that is)?

Here is one sketch. Note first that it would seem to be wholly arbitrary and inexplicable if there were just, say, two minds in existence associated with physiologically functioning human qualitative profiles, but billions of other non-mined humans with similar physiological profiles. We could thus run an explicability argument, perhaps.

This arbitrariness/explicability worry, moreover, is magnified if we accept a *principle of relevant differences* according to which, roughly, if x and y only differ in respects that are irrelevant with regard to possessing some further property P (or having some further fact F true of them), then it is inexplicable (else: metaphysically impossible) for one of x or y to have P (or F) but the other to lack P (or F). From this, perhaps we could argue that the only differences that obtain between the Q_i profiles seem quite clearly to be irrelevant with respect to having H_i be true of them. For instance, surely the mere *height* of something makes no difference with respect to that individual's possessing a mind – similarly for age, skin color, hair color, eye color, slightly different behavioral tendencies, and so on. But such irrelevant differences seem to be the only differences one could point to in our scenario concerning minded versus purportedly non-minded humans.

Second, surely the feature 'having a mind' is an essential feature of a thing. A mind (or at least a disposition towards having or developing a mental life, as in the case of human fetuses, people in a coma, people in deep sleep, and so on) seems to be the sort of thing that is built into the very nature of a thing which has it and is not just some contingent accident that something happens to possess. But if that is the case, then if at least two humans have minds, it follows that all humans (*qua human*) must have minds.

Admittedly, this pushes the problem back a step insofar as we now face the problem of justifying why all the other seemingly human individuals with profiles Q_1 through Q_n *do*, in fact, share a common nature with the two humans we have established as minded (as opposed to, say, having the nature of human*, where the nature of human* consists in sharing nearly

identical qualitative non-minded profiles with humans but essentially lacking minds).

Nevertheless, absent any defeaters to the contrary, surely we are warranted in defeasibly inferring that two things share a nature in common when they are nearly identical in terms of all of their non-mental properties (like physical constitution, evolutionary history/origin, characteristic behavioral tendencies, nearly identical causal powers, and so on). From this, we could perhaps mount a defeasible argument for all humans' having minds.

Perhaps we could also motivate the inference to all humans' being minded by appeal to a principle of *modal uniformity* (Rasmussen (2014)). In particular, we may hold that because (i) humans are categorically alike, and (ii) the only differences that obtain between humans are degreed differences, we are justified in inferring that such degreed differences do not make a difference with respect to modal properties like the contingency of humans, their possessing minds (which is a modal property provided being minded is an essential feature of a thing), and so on.

Finally, perhaps we could give two epistemological solutions to build off of Pruss' argument. The first makes use of a thesis of phenomenal conservatism while the second makes use of externalism. Roughly, the thesis of phenomenal conservatism I have in mind is as follows, where S is a subject and P is a proposition:

> If it seems to S that P is true, then S has *prima facie*, defeasible evidence for P's being true.

So, perhaps we could argue that if it seems to one that other humans are minded, then absent any defeaters, one has evidence in favor of other humans' being minded. And perhaps this is bolstered by Pruss' argument, given that we know it is not only *possibly* true, but also *actually* true that there exists at least one human apart from ourselves that is minded.

Finally, one may adduce externalism as an extension of Pruss' argument. Externalism is the thesis that, roughly, factors external to the knower confer justification to the knower's beliefs. The nature of justification, in other

words, is not completely determined by internal factors alone. One form of external justification is standing in an appropriate causal connection to the fact known.

Crucially, though, suppose it is in fact the case that all humans are minded. Then, presumably, *their being minded* is what causes their seemingly intelligent and intentional behavior. And their seemingly intelligent and intentional behavior is, in turn, precisely what causes or induces us to believe that they are minded. So, there seems to be an appropriate causal connection between the *fact* that humans are minded and our *belief* that humans are minded (on the supposition that humans are in fact minded). It seems, then, that if humans are *in fact* minded, then our belief that they are minded counts as knowledge under externalist accounts of knowledge (provided that other conditions are met of course, like reliably functioning cognitive faculties, etc.).

If this is true, then it seems that whether or not we know humans are minded depends primarily on the *actual fact of the matter*, in which case arguments attempting to demonstrate that we do not know other humans have minds must *establish that it is false that other humans have minds* – not merely that we have no specific reason to rule out certain hypothetical, remote epistemic possibilities of (say) philosophical zombies.

Combining this with Pruss' argument, we can see that no such skeptical argument could demonstrate the more general thesis that *no other human has a mind*, since (per Pruss' argument) we are virtually certain that this is false. The skeptic of other minds must therefore argue for a more restricted and limited thesis, while avoiding his reasons for the restricted and limited thesis equally applying to the known-to-be-false thesis that *no other human has a mind*. This certainly seems to place the skeptic on the back foot. Moreover, Pruss' argument shows that it is certainly *possible* to stand in such an appropriate causal connection with another mind (given our virtual certainty that another such mind actually exists). Thus, there is no barrier *in principle* to attaining such knowledge, contrary to what many proponents of skeptical arguments claim.

7 Conclusion

We've come quite a long way on our journey, gaining fresh insights into the nature of the mind. We first examined Descartes' substance dualism, after which we evaluated Armstrong's identity theory. After an in depth investigation into behaviorism, we finished with a critical examination of eliminativism and a potentially groundbreaking argument for the existence of other minds.

The investigation, though, is never over. Reason continues to reveal insights the further and deeper you tread. Think of this chapter, then, as an invitation into the depths of reason.

Chapter 6: Conclusion

We commenced our journey with a consideration of tribalism and its pernicious effects across a range of domains. We then sought antidotes to tribalism, two of which became prominent: *reason and love*. With reason in hand, we can reduce the blinding effects of us-versus-them bias and focus on the issues and arguments themselves. With love in hand, we can see our interlocutors as human beings with value, worth, and dignity.

We also considered various intellectual virtues as well as twenty-two tips for the productive exchange of ideas. These laid the foundation for the rest of our journey.

Next, we turned to methods of argument evaluation and a variety of strategies and tools for your philosophical toolkit. We then applied these in three different contexts. The first context was an analysis of scientism, the view that science is the only source of knowledge. The second context was laws of nature, in which we found a problem of persistence for Humean Regularity Theory. The third and final context was the nature of the mind. We explored each of these contexts using the skills, methods, strategies, and tools developed in earlier parts of the book.

But we also saw first-hand the beauty and power of reason in revealing truths that *serve*. Don't let the intricacies of arguments detract from your appreciation of this point.

As our journey *together* nears an end, the journey of truth-seeking has just begun. Dare to question. Dare to think. Liberate yourself from the chains of tribalism. Challenge straw with steel. Reduce your chains to naught.

Ridding yourself of chains takes courage and effort. The bondage of tribalism is weighty. Perhaps, though, our inherent tendencies toward us-versus-them thinking create the ideal crucible in which we are transformed.

Perhaps *loving* those you deeply wish to combat is one of the greatest ways to cultivate virtue and character. Perhaps it is this endless struggle that creates a love for the immortality of reason.

The journey is now yours. Only you can determine your path forward. Where will reason take you? What hills and valleys will you trek, what treasures will you seek? Amidst this uncertainty, at least one thing is certain: reality reveals its wonders to patient, inquisitive seekers. And if you start to lose hope on your journey, remember one thing: the Majesty of Reason is always by your side.

Bibliography

Armstrong, D. (1980). The Nature of Mind. University of Queensland Press.

Augros, M. (2017). The Immortal in You: How Human Nature is More Than Science Can Say. Ignatius Press.

Aquinas, T. Summa Theologiae, ST I q2,a3.

Ayer, A.J. (1936). Language, Truth and Logic. London: Victor Gollancz Ltd.

Baker, K. (2014, March 20). Philosophy of Mind 2 – Logical Behaviorism. [Video file]. Kane B. Retrieved from

https://www.youtube.com/watch?v=lwat9ZFpHNU

Cappelen, H. (2017). Disagreement in Philosophy: An Optimistic Perspective. In G. D'Oro, & S. Overgaard (Eds.), Cambridge Companion to Philosophical Methodology.

Churchland, P. (1984). Matter and Consciousness. MIT Press.

Dowden, B. Fallacies. Internet Encyclopedia of Philosophy.

Elder, L. (2007, September). Retrieved from "Our Concept and Definition of Critical Thinking," The Foundation for Critical Thinking, https://www.criticalthinking.org/pages/our-conception-of-critical-thinking/411

Feser, E. (2017). Five Proofs of God's Existence. Ignatius Press.

Feser, E. (2019). Aristotle's Revenge: The Metaphysical Foundations of Physical and Biological Science. Editiones Scholasticae.

Hulpke, A. (2015). Basic Argument Forms. Mathematics 192, First Year Seminar. Retrieved from https://www.math.colostate.edu/~hulpke/lectures/m192/proofmodes.pdf

Lowder, J. (2018). Opening Statement from my Debate with Frank Turek. The Secular Outpost, https://www.patheos.com/blogs/secularoutpost/2018/12/29/opening-statement-from-my-debate-with-frank-turek/#_ftn6

Malpass, A. & Morriston, W. (2020). Endless and Infinite. Philosophical Quarterly. (forthcoming)

Morrow, D. & Weston, A. (2015). A Workbook for Arguments 2nd ed. Hackett Publishing Company.

Novela, S. (2018). The Skeptic's Guide to the Universe: How to Know What's Really Real in a World Increasingly Full of Fake. Grand Central

Publishing.

Papineau, D. (2012). Philosophical Devices: Proofs, Probabilities, and Sets. Oxford University Press.

Paul, L.A. & N. Hall. (2013). Causation: A User's Guide. Oxford University Press.

Pruss, A. & Rasmussen, J. (2018). Necessary Existence. Oxford University Press.

Pruss, A. (2006). The Principle of Sufficient Reason: A Reassessment. Cambridge University Press.

Pruss, A. (2019, September 19). There are other minds. Blogspot. Retrieved from http://alexanderpruss.blogspot.com/2019/09/there-are-other-minds.html

Rasmussen, J. (2014). Continuity as a Guide to Possibility. Australasian Journal of Philosophy, 92(3), pp. 525-538.

Rasmussen, J. (2015). Building thoughts from dust: a Cantorian puzzle. Synthese, 192, pp. 393-404.

Rasmussen, J. (2017, April 12). A closer look at why things continue to exist (episode 8 of 20). [Video file]. Worldview Design. Retrieved from https://www.youtube.com/watch?v=_GyNcuI4p6c

Rasmussen, J. (2019, July 17). Can you know anything in philosophy? [Video file]. Worldview Design. Retrieved from https://www.youtube.com/watch?v=gC7Vkp3Km8E

Rosenberg, A. (2011). The Atheist's Guide to Reality: Enjoying Life without Illusions. W.W. Norton & Company.

Russell, B. (1945). A History of Western Philosophy. Allen & Unwin.

Ryle, G. (1949). The Concept of Mind. Chicago University Press.

Sanford, D. (1999). Circular Reasoning. In R. Audi (Ed.), The Cambridge Dictionary of Philosophy, 2nd ed. Cambridge University Press.

Scriven, M. & Paul, R. (1987). Annual International Conference on Critical Thinking and Education Reform. Retrieved from the Foundation for Critical Thinking, https://www.criticalthinking.org/pages/defining-critical-thinking/766

Speaks, J. (2011, September 27). Verificationism. PHIL 83104 Handout. Retrieved from https://www3.nd.edu/~jspeaks/courses/2011-12/83104/handouts/verificationism.pdf

"Waterfall diagrams and relative odds." Arbital. Retrieved from https://arbital.com/p/bayes_waterfall_diagram/?l=1x1

Whately, R. (1826). Elements of Logic. Boston and Cambridge: James Munroe and Company.

---

[1] To those of you who are grammar savants: I tried very hard to re-write this sentence so that it didn't end in a preposition. You're welcome. (Though, I cannot guarantee grammatical perfection elsewhere. As the distinguished intellectual Miley Cyrus has emphasized, *nobody's perfect*).

[2] These are in no particular order.

[3] Or 'centre-stage', if you live in a place with a cool accent.

[4] 'Straw-manning' refers to constructing (and criticizing) a *weak* form of someone else's argument. 'Steel-manning' is the opposite; it refers to intentionally constructing an *improved* version of someone else's argument.

[5] Don't worry about the word 'deductive' – we'll cover that soon.

[6] Once again, don't worry too much about formal logic (at present) – we will briefly cover that in due time.

[7] Though, technically, it depends on the dialectical context at hand. Let's just suppose we are in a dialectical context wherein each party disagrees about whether free will exists – a libertarian (in the philosophical sense) and a hard determinist.

[8] But, once again, this will depend on the particular dialectical context at hand. Some people may not realize that determinism entails that we cannot justifiably praise and blame others, and because (say) such people already accept that we can justifiably praise and blame others, this argument may give them reason to reject determinism. Once again, this illustrates how whether or not an argument is question-begging depends on the dialectical context at hand. Don't get too caught up into the details of the argument against determinism – it is simply an illustration to help you understand circular reasoning (question-begging).

[9] A philosophical zombie is a being that is physiologically indistinguishable from a human but that has no mental states (like pain, fear, thoughts, desires, and experiences).

[10] Once again, this is not unqualifiedly true – whether or not a question is begged depends on the dialectical context. In some contexts, philosophers give *arguments* or *reasons* to think zombies are possible (for instance, that they are conceivable, coherently imaginable, internally consistent, and so on). We can stipulate, however, that we are in a context wherein one levels this argument without mounting independent arguments for premise (1).

[11] Assuming, once again, that independent motivation is not provided for premise (1).

[12] Once again, these are in no particular order. Also, just because I provide examples wherein an argument for position P is criticized, that neither means nor entails that I deny P (nor does it mean or entail that I accept P). All the examples are used for *illustrative* purposes only.

[13] 'Proof' and its cognates should really be used sparingly (see tip #13). I use it in this section because the oft-repeated phrase in contemporary discourse is 'you cannot prove a negative'.

[14] And do you know what we've just done? We have proven that *it is not the case that you cannot prove a negative*… meaning we have just proved another negative!

[15] A 'token' is a *particular instance* of a general or universal 'type' or kind. Consider this list: dog, dog, dog. How many words are in the list? Well, there are three word *tokens* (particular instances of the type of word 'dog'), but there is only one *type* of word.

[16] Philosopher Paul Draper calls this the *fallacy of understated evidence*.

[17] Without this presupposition, pain wouldn't play a biological role, since no natural disasters, predators, etc. would *harm* animals so as to cause (i) tissue damage and (ii) the need for self-preserving behaviors.

[18] Once again, I am not taking a stance on the problem of animal suffering or on the hypothesis evaluated in this example. These are for illustrative purposes only.

[19] For instance, defenders of eternalism – the view according to which past, present, and future all actually exist – appeal to various scientific experiments central to the General Theory of Relativity.

[20] Epistemology is the study of knowledge whereas metaphysics is the study of the fundamental nature of reality.

[21] Hesperus and Phosphorus are identical to the planet Venus.

[22] This refers to metaphysical possibility. X is metaphysically possible if and only if X exists at least one possible world. In other words, X *can* exist.

[23] A possible world is just a complete way that the world could have been (world means not just the universe but *all of reality*).

[24] This follows from the S5 axiom of modal logic.

[25] At least if we accept S5 modal logic. For a defense of S5, see Pruss & Rasmussen (2018).

[26] This refers to metaphysical possibility. X is metaphysically possible if and only if X exists at least one possible world. In other words, X *can* exist.

[27] A possible world is just a complete way that the world could have been (world means not just the universe but *all of reality*).

[28] This follows from the S5 axiom of modal logic.

[29] Perhaps not.

[30] Though, there are different senses of 'natural'. One is just 'found in the natural, physical world and not human-made'. Another is 'facilitates the flourishing of human nature'. Let's restrict ourselves to the former meaning for now.

[31] Never mind why there is a cow in the boat.

[32] Utter-ly. Cows. Ha!

[33] Points for persistence, at least @Jones.

[34] Intentionality refers to the mind's 'of-ness', 'about-ness', or 'directed-ness' toward certain objects. In other words, it refers to the mind's ability to *represent* and *mean* things.

[35] The examples are adapted from Papineau (2012).

[36] Assuming he has demonstrated the impossibility of infinitely long chains of changes, of course.

[37] I am not saying Aquinas' arguments concern chains extending infinitely in the past. On the contrary, such chains constitute *per accidens* chains, whereas Aquinas concerns himself with *per se* chains. Don't worry if you don't know these terms – I delegated this to a footnote for a reason.

[38] For a defense of this principle, known as the Principle of Sufficient Reason (PSR), see Pruss (2006).

[39] The following discussion is based on Lowder (2018) and on Paul Draper's work in Bayesianism.

[40] It's difficult to give a precise account of probability. Unless stated otherwise, I mean 'epistemic probability,' i.e. a rational degree of credence or expectation.

[41] Though, pay attention – we'll use this later.

[42] The example is adapted from "Waterfall diagrams and relative odds".

[43] Goldbach's conjecture is a (ye) unproven mathematical conjecture stating that every even integer greater than two is the sum of two prime numbers.

[44] Here is a small technical complication. A more accurate way to capture this inference's reliability (that takes into account Bayesian conceptions of evidence and probabilistic updating) is as follows: Not seeing P's truth (or S's obtainment) is evidence of P's falsity (or S's non-obtainment) only if *were* P to be true (or S to obtain), it is *more likely* that conditions (i) and (ii) are jointly met than were P to be false (or S not to obtain), where conditions (i) and (ii) are as follows: (i) we are aware that P is true (or that S obtains), and (ii) we recognize P (or S) *qua* P (or *qua* S).

[45] That is, the thing in reality – 'out there' – to which the expression refers.

[46] Sentence translations at the end of each rule (see below) are thanks to Hulpke (2015).

[47] Take note – this is the construction of a parallel argument that demands a symmetry breaker!

[48] Or, at least, a successful probabilistic argument that utilizes resource compatible with verificationism. If we didn't have this relevant *background rate* of the proportion of universally true laws out of the reference class of natural regularities, then the only way we could justify P is by appeal to something *other than* observational frequency – and that's disallowed on verificationism.

[49] Or, as we phrased it in the previous chapter, 'kicks the can back a step'.

[50] Modus ponens is the logical rule of inference according to which *if p entails q and p is true, then q is true*.

[51] The cunning among you will recognize this distinction from the first tool in the previous chapter.

[52] The live possibility given only the methods and tools of *science*, that is.

[53] A reductio (*reductio ad absurdum* is the full phrase, but very few use it), in this context, is understood simply as a proof by contradiction (assume Q, show a contradiction follows from Q, and then reject the initial assumption, i.e. conclude ~Q).

[54] The argumentative structure is really as follows. We know that if P is morally responsible, then APC and SC are met for P. Let 'P is free' be 'Q' and 'APC and SC are met' be 'R'. So, (i) we know that Q entails R. But (ii) if we know that Q entails R, and we are justified in believing Q, then we are justified in believing that R. But (iii) if scientism is true, we cannot be justified in believing that R. From (i), (ii), and (iii), it follows that if scientism is true, we cannot be justified in believing that Q. But if we cannot be justified in believing that Q, then we cannot be justified in believing anyone is morally responsible. But we *can* be justified in believing that someone is morally reponsible. So, scientism is false.

[55] Constructive dilemma: P entails Q; and R entails S; but either P or R; so, either Q or S.

[56] Informally, the principle of relevant differences just says that in situations wherein parallelism is present (say, exactly parallel arguments), there must be some symmetry breaker (i.e. relevant difference between the parallel arguments or propositions) if one parallel argument/proposition is true or justified at the expense of the other parallel arguments/propositions.

[57] One of the causes, rather.

[58] There are different versions of regularity theory, some more sophisticated than others. Arguably, though, the one examined here is the one Hume's commitments seem to lead to.

[59] If this sounds eerily familiar, your sense is correct. Here is Hume himself expounding this view: "If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, Does it contain any abstract reasoning concerning quantity or number? No. Does it contain any experimental reasoning concerning matter of fact and existence? No. Commit it then to the flames: for it can contain nothing but sophistry and illusion. (Hume, *An Enquiry Concerning Human Understanding*)

[60] Keep in mind that probabilities (in this context) refer to epistemic probabilities. Epistemic probabilities are understood as *reasonable degrees of expectation, belief, or credence*. In a probability analysis which employs epistemic probabilities, we are concerned with the degree to which we expect an outcome given our justified acceptances. Imagine tossing a fair coin. Our rational degree of expectation ought to be 0.5 for its coming up heads. We will give heads a 0.5 chance even if the world is deterministic, i.e. even if the coin was determined to come up heads by the nature of the toss and the laws of physics (say). The point is that you have no idea which side it will come up, and thus your rational degree of expectation is (and ought to be) 0.5.

[61] We actually have good reason to think that the law-like regularities have persisted for billions of years, but I'll stick with the much more conservative number.

[62] The principle of indifference states that, absent any considerations favoring on outcome over the others, distribute your credences equally among the alternative outcomes. In this case, there are two outcomes: (i) persistence in accordance with lawful regularity at s or (ii) no persistence in

accordance with lawful regularity at s (i.e. a 'violation' of the thus-far-established regularity). Because HRT gives us no reason to expect either one of (i) or (ii) over the other, the principle of indifference advises us to give each outcome a 0.5 probability given HRT.

[63] That is, what happens at one second doesn't raise or alter our expectations of what happen at another, later second. HRT requires this, since there aren't any grounds or explanations of persistence, and *a fortiori* there are no explanations of persistence in terms of things that happened at earlier seconds. Hence, what occurs at previous seconds doesn't alter our expectations of what occurs at later seconds with respect to the likelihood ratio.

[64] Though, all the argument needs to succeed is that it's a little greater than 0.5 (say, 0.6 or even 0.51) – the math would come out approximately the same either way.

[65] There are some technical complications, such as a potential intermediate view that holds that, while *Persistence* is not necessitated, it is made probable by (say) the causal powers and dispositions of things. But we can set these aside, since (i) these do not ultimately bear on the argument in question; (ii) these are incompatible with Humean Regularity Theory; (iii) a similar argument can be mounted in favor of a necessitating account (~HRT) against any account which makes the conditional probability of persistence in accordance with lawful regularities at any given second s less than 1; and (iv) even if these make ~HRT less intrinsically probable than HRT, they certainly won't do it to a degree required to overturn the massive discrepancy in the likelihood ratio.

[66] Indeed, Rasmussen (2017) was the inspiration for this problem for Humeanism. While I've adapted and innovated it, the basic idea stems from Rasmussen (2017).

[67] Again, we would need some symmetry breaker between the conceivable identity between mind and brain (on the one hand) and conceivable separability of mind and brain (on the other).

[68] The extension of some expression or property is just the set of things in reality that 'fall under' or 'answer to' that expression. Two sets are co-extensional when the same individuals that fall under one expression fall under the other as well. For instance, 'creature with a heart' and 'creature

with a kidney' are co-extensional, as are 'having a radius' and 'being a circle'.

[69] Again, don't worry if this sounds foreign to you. This is a bonus argument.

[70] Or, perhaps they are possessed by a conscious subject, but the subject flickers in and out of existence with the thought. So, there would be no singular, unified subject of multiple conscious states. Also, if this sounds odd, it's because of the incredibly high bar Descartes places on knowledge. For Descartes, knowledge is certain and indubitable cognition.

[71] Those strategies and tools in chapter two have far greater applicability than you may have initially thought!

[72] The problem of other minds consists in the following question: how do we know that other people have subjective, first-person, minded consciousness?

[73] Once again, a *symmetry breaker* is needed.

[74] I won't defend this particular claim at length here, though it seems quite plausible to me. It captures the idea behind the general Peripatetic axiom that *nihil est in intellectu quod non prius fuerit in sensu* – nothing is in the intellect that was not previously in the senses. Even if one disagrees with my claim in the text, we still get a deeply significant result: *if* it's the case that our concepts can never go beyond what we have encountered in sensory experience except by means of negation, combination, or abstraction, it follows that *behaviorism is false*.

[75] Suppose that the concept of private, subjective mental states is nothing more than a mere combination of the concepts of public, objective things. If that is true, then it seems that we can analyze our concept(s) of the privacy and subjectivity of mental states in terms of the concept(s) of public and objective realities. But it seems we cannot analyze our concept of the private and subjective nature of mental states in terms of concepts of purely public and objective realities, since any such analysis will necessarily leave out the very thing needing to be analyzed, namely subjectivity. When given

an analysis of (our concept of) private, subjective mental states solely in terms of (our concept of) public and objective facts, it seems there will always be the further question: *why* are those objective features accompanied by this rich inner experience, a what-it-is-like-ness, a private and subjective 'feel'? And why are these objective facts linked with *those specific subjective features* rather than others or no subjective features at all? This gap will persist by the very nature of the proposed analysis, it seems. This seems to be a second reason against the combination view.

[76] But that's a *good* thing!

[77] Unless you're an eliminativist, since an argument can't invite further thought if thoughts don't exist.